Planetary Giant Impacts: Simulating Collisions and their Consequences

KEGERREIS, JACOB,ALEXANDER

Planetary Giant Impacts

Simulating Collisions and their Consequences

Jacob Kegerreis

A thesis presented for the degree of
Doctor of Philosophy

Institute for Computational Cosmology
Department of Physics
Durham University
United Kingdom

November 2019

Planetary Giant Impacts

Simulating Collisions and their Consequences

Jacob Kegerreis

Abstract:

Giant impacts dominate many planets' late accretion and evolution, but the detailed consequences of these violent events are still poorly understood. In this thesis I use 3D smoothed particle hydrodynamics simulations of giant impacts to study three primary topics: the impact origin of Uranus' obliquity; numerical convergence with increasing resolution; and atmospheric erosion by giant impacts.

To these ends, the SWIFT code is developed to model planetary impacts with 1000× more simulation particles than the current standard, and an efficient method for creating relaxed initial conditions is presented. A suite of simulations of giant impacts onto the young Uranus confirms that the planet's high obliquity can be explained by a wide range of impact scenarios. For some grazing collisions, most of the impactor's ice and energy is deposited in a thin shell in the target's outer ice layer, which might help explain Uranus' observed lack of heat flow from the interior. Follow-up simulations with just over 10^8 simulation particles reveal that standard simulations with fewer than 10^7 particles fail to converge on even bulk properties like the post-impact rotation period, or on the detailed erosion of the atmosphere. Higher resolutions appear to determine these large-scale results reliably, but even 10^8 particles may not be sufficient to study the detailed composition of scattered debris. This improvement in resolution then enables the first full, 3D simulations of atmospheric erosion on terrestrial planets by giant impacts. For head-on collisions, there is a rapid change with increasing impact speed from very little erosion to total loss. However, for grazing impacts there is a more gradual change with speed and a non-monotonic dependence on the impact angle. These projects highlight the necessity of high-resolution simulations in three dimensions to capture the complexity of giant impacts.

In the final chapter, a model of the Moon's argon exosphere is used to test competing explanations for the strange features observed by the recent LADEE mission. The persistent overdensity of argon over the maria can only be reproduced by a localised endogenic source. This offers a novel probe of the lunar interior late after its origin in a giant impact.

Contents

List of Figures

List of Tables

Flipbook Animations

In the bottom-right corner, starting from the final page (133), are tiny snapshots of a grazing giant impact onto a proto-Earth, as described in Chapter 4, with one frame every 1,400 s. From page 67, a more head-on collision with the young Uranus is shown, as described in Chapter 3, with snapshots every 500 s.

Declaration

The work described in this thesis was undertaken between between October 2015 and August 2019 while the author was a research student under the supervision of Dr Vincent Eke and Prof Richard Massey in the Department of Physics at Durham University. No part of this thesis has been submitted elsewhere for any other degree or qualification.

The contents of Chapters 2, 3, and 5 have been published in the following papers:

- *Planetary Giant Impacts: Convergence of High-Resolution Simulations using Efficient Spherical Initial Conditions and SWIFT*
  Jacob Kegerreis, Vincent Eke, Pedro Gonnet, Donald Korycansky, Richard Massey, Matthieu Schaller, Luís Teodoro.
  Mon. Not. R. Astron. Soc. 487:4, 2019. doi: 10.1093/mnras/stz1606

- *Consequences of Giant Impacts on Early Uranus for Rotation, Internal Structure, Debris, and Atmospheric Erosion*
  Jacob Kegerreis, Luís Teodoro, Vincent Eke, Richard Massey, David Catling, Chris Fryer, Donald Korycansky, Mike Warren, Kevin Zahnle.
  Astrophys. J., 861:52, 2018. doi: 10.3847/1538-4357/aac725

- *Evidence for a Localized Source of the Argon in the Lunar Exosphere*
  Jacob Kegerreis, Vincent Eke, Richard Massey, Simon Beaumont, Rick Elphic, Luís Teodoro.
  J. Geophys. R. (Planets), 122:2163–2181, 2017. doi: 10.1002/2017JE005352

All figures in this thesis were produced by the author, or have been properly attributed in the figure caption.

Acknowledgements

I am immensely grateful to my supervisors, Vince Eke and Richard Massey, for supporting, inspiring, and guiding me throughout my PhD. Thank you for sharing so much of your time, enthusiasm, and invaluable advice; for helping me to work through stubborn problems and reminding me to step back to see the bigger picture; for motivating me with kinder encouragement than I felt I deserved and sharing the excitement when we made progress.

Huge thanks also to Lydia Heck and the COSMA team for being so generous with their intimidating expertise and equally eminent friendliness. Thanks to Luís Teodoro and the now-many external colleagues I've had the pleasure to meet for their inordinate support and encouragement. Thanks to Matthieu Schaller and the SWIFT team for being so willing to fix things I break while helping me learn to fix them myself next time.

I feel so lucky to have had the company of great course-mates, office-mates, colleagues, and friends – far too many to try to list everybody here without forgetting someone, so I won't. Thank you for making Durham an enjoyable place to be, in and outside of work, with special thanks to my longtime housemates Phil, Jack, and Kris. Thanks to all the above and to Sarah, my extended family, and family friends for your support and for always acting interested in hearing about my research.

Finally, I owe more heartfelt thanks to my parents than I will ever be able to put into words. Plus I've used too many superlatives in this text already. Thank you for everything.

This thesis was supported by STFC grant ST/N50404X/1 and the ICC PhD Scholarships Fund.

Let's work the problem, people. Let's not make things any worse by guessing.

—*Gene Kranz, Apollo 13*

What are you talking about? My god, pull yourself together. 'What will you do?' Is, is, is this a question? Go, confront the problem! Fight. Win!

—*Edna Mode, The Incredibles*

Dedicated to

Mum and Dad

Chapter 1

Introduction

GIANT impacts play a crucial role in planet formation, building planets out of colliding embryos in the early solar system and provoking critical developments in their later evolution (Chambers, 2001). These violent, messy, and highly non-linear processes are most commonly studied using smoothed particle hydrodynamics (SPH) simulations.

This thesis focuses on three primary questions about giant impacts that have long been overdue attention and progress:

1. Can a collision explain Uranus' obliquity, and what are the other consequences for the planet?
2. Are the results of standard SPH simulations numerically converged, and what can we learn from higher resolution studies?
3. How much atmosphere is ejected by a giant impact?

To tackle the numerical challenge of these simulations, I first develop the hydrodynamics code SWIFT to model planetary impacts with 1000× more simulation particles than the current standard, and create an efficient method for generating relaxed initial conditions.

I also include a project on modelling the Moon's argon exosphere, which, with its implications for the interior of our satellite, provides a potential present-day link to the Moon's origin following a giant impact.

The projects in this thesis deal with some fairly distinct topics, so the specific background and motivations for each of them are discussed in more detail at the start of each chapter. In this general introduction, I briefly review the importance and the history of giant impact research, the bigger picture of planet formation, and a small subset of seminal papers. The remaining chapters of this thesis are then outlined in §1.3.

1.1 Giant Impacts and SPH Simulations

Cataclysmic collisions have long been known to dominate the late stages of terrestrial planet formation. A significant proportion of the work in the field has focused on the origin of our Moon, perhaps unsurprisingly, given the far-superior observations and samples compared with other systems and the cultural interest in the question. Furthermore, the Moon's large size, small iron core, and relative lack of water all fit readily with the idea of a planet-fragmenting, volatile-depleting collision. That being said, the earliest study of a giant impact affecting a planet in our solar system was Safronov (1966). They estimated the sizes of the largest accreting planetesimals and showed that Uranus' remarkable tilted rotation could be caused by an impactor with around 5% of the ice giant's mass.

If we want to look even further back, then the underlying idea that the early Solar System hosted many large planetesimals in the first place was solidified by Urey (1952, 1959) (and refined in the context of giant impacts by e.g. Wetherill, 1976). Urey suggested that tens of Moon-sized objects formed and were broken up in collisions to create the early meteorites, in order to explain their composition of fragmented metal aggregates and post-formation mixes of silicates with metals.

Hartmann & Davis (1975) further developed this idea and found it likely that planetesimals approaching the size of Mars would form to collide with each other and the young planets. They claimed that more than double the Moon's mass of material could be ejected from the Earth by a Mars-sized impactor. They also pointed out the 'philosophically satisfying' idea that the strange features of the different planets are caused by the small-number statistics inherent in the details and fate of the largest secondary planetesimal that forms in each system.

This impact hypothesis for the origin of the Moon was strengthened by Cameron & Ward (1976), who focused on the high angular momentum of the resulting system and the trajectory of debris. For a suitably large tangential collision that could explain the angular momentum, they argued for the rapid accretion of a suitably iron- and volatile-poor Moon from the disk. This growing theoretical confidence was bolstered by the numerical simulations of Cameron (1985), who modelled the evolution of test particles in a simplified collision to show how more than a lunar mass could end up in orbit.

Benz et al. (1986) then performed the most direct predecessor simulations for the projects in the field and in this thesis. Using smoothed particle hydrodynamics (SPH) – which had been first developed to study star formation (Lucy, 1977; Gingold & Monaghan, 1977) – Benz et al. ran full, 3D hydrodynamical models of Moon-forming

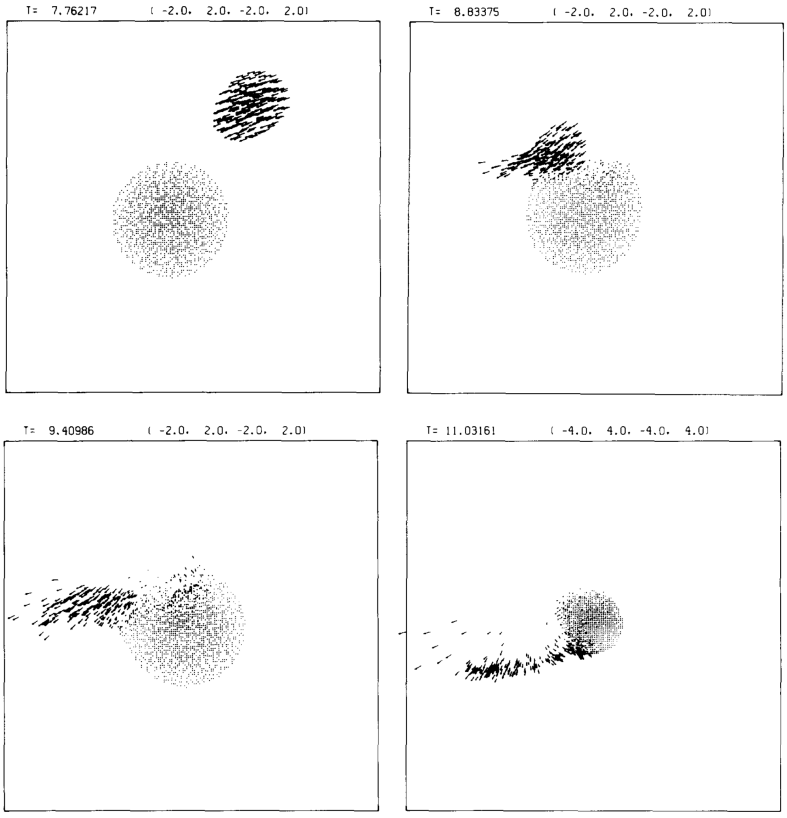


Figure 1.1: Snapshots from the first simulations of a giant impact, by Benz et al. (1986, Fig. 2) who used $\sim 10^3$ SPH particles to model the creation of a Moon-forming debris disk. Arrows show the particles' velocities. The time in hours and the axis limits in Earth radii are shown at the top of each frame.

giant impacts. Fig. 1.1 shows snapshots from one of their simulations with 1,216 particles. This also highlights the computational advances over the last three decades that have made possible the >100,000,000 particles of the SPH simulations in this thesis.

Shortly thereafter, we are brought back to the first giant impact subject with Slattery et al. (1992)'s SPH simulations of Uranus impacts, now up to 8,000 particles as shown in Fig. 1.2. They confirmed that a wide variety of collisions could produce the tilted spin we see today. Unlike the formation of the Moon, the origin of Uranus' obliquity has been investigated little since these early simulations.

Following these pioneering papers, many different types of giant impacts have been hypothesised and simulated (often using SPH) as explanations for weird features on almost every planet in the Solar System. In addition to advances in understanding of the planets' pre- and post-impact evolution, these developments continue to spur on and be guided by improvements in numerical simulations: both in terms of the hardware and availability of supercomputers, and the software with more

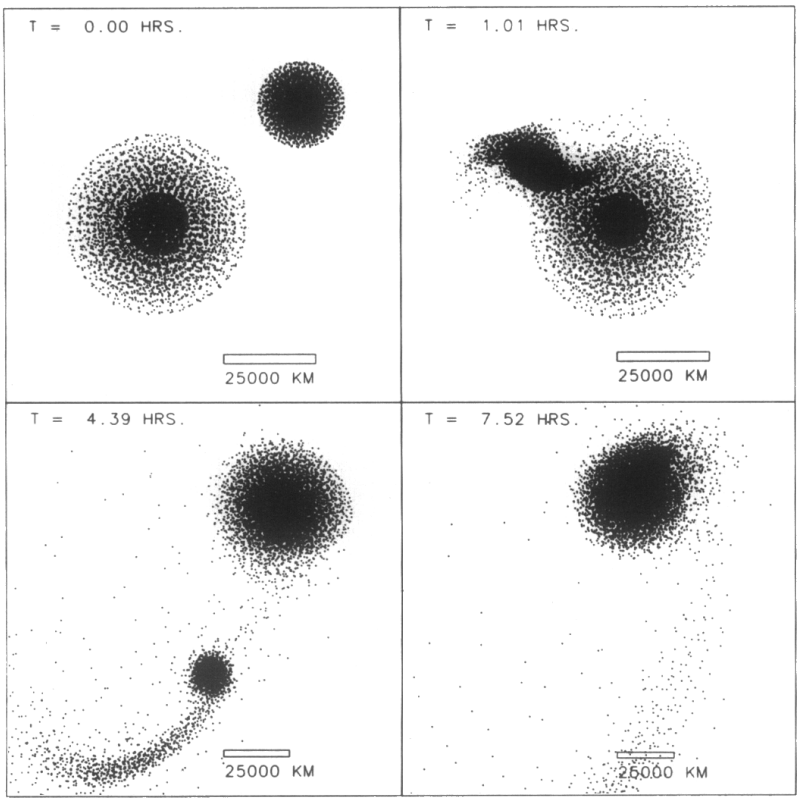


Figure 1.2: Snapshots from Slattery et al. (1992, Fig. 2)'s simulations of a giant impact on Uranus using $\sim 10^4$ SPH particles.

sophisticated and purpose-built codes.

For example, Mercury's thin rocky mantle is thought to be the result of a violent impact that blasted off the outer layers (e.g. Benz et al., 2007). Slightly less large collisions have been blamed for both Mars' moons and its 5 km-lower northern hemisphere, perhaps even from a single event (e.g. Nimmo et al., 2008; Citron et al., 2015). Smaller impacts again could produce some asteroid families, like Vesta (e.g. Asphaug, 1997). Neptune, similar to Uranus in mass and composition, might have had its interior homogenised by a head-on, penetrating giant impact compared with a grazing, spin-inducing collision on Uranus (e.g. Podolak & Helled, 2012).

One of the challenges of studying these problems is the huge parameter space of possibilities. Even as models of early planet formation improve, the initial conditions for a collision remain acutely uncertain – not just the obviously important speed and angle of the impact, but the mass, composition, temperature, and internal structure of both the target and impactor also matter, with a wide range of plausible values. Furthermore, we only see the planets billions of years after the giant impacts that we want to study. So not only are the initial conditions unknown, but the outcomes require a daunting amount of post-processing and extrapolation in order to compare them with observations.

For example, in the case of the Moon's formation, SPH simulations can produce a debris disk. Then, disk-evolution and accretion models provide estimates for the early Moon's mass, composition, and orbit. Finally, long-term orbital and tidal models can test whether the satellite would evolve towards the system we see today. More and different techniques are required to answer different questions, such as where does any ejected debris go, how is the Earth's geochemical evolution affected, or what happens to a pre-impact atmosphere. This makes giant impacts a topic that brings together many different models and disciplines to study in its entirety. That said, most projects, including those in this thesis, tend to focus on the collisions themselves and their immediate aftermath. There is plenty that we do not understand about what happens during the tens of hours of a giant impact, let alone in the following billion years.

The state of the art for these studies is still dominated by SPH simulations, typically using 10^5 up to 10^6 particles at the time of writing. Grid-based codes with comparable spatial resolutions have also been applied to model giant impacts, but are more commonly used for smaller scale impacts and 2D simulations (e.g. Wada et al., 2006; Collins, 2014). In context, this number of SPH particles means that in a standard simulation of a Moon-forming impact, the debris disk that is inspected for forming the Moon would be practically unresolved at only ∼5 particles thick. Such simulations therefore cannot reliably predict the composition, structure, or thermodynamic state of the disk. This leaves significant room for improvement.

1.2 Planet Formation and Exoplanets

Stepping back from the present-day quirks of the planets in our neighbourhood, giant impacts are of course important for the evolution of planets in the first place. Terrestrial planets typically form from tens of roughly Mars-sized planetary embryos that crash into each other after accreting from the proto-planetary disk. Two of the seminal ideas for the overall formation of our solar system are the Nice and Grand Tack models. Both focus on the evolution of the giant planets, but, unsurprisingly, this impacts the rocky planets as well.

The Grand Tack model (Walsh et al., 2011) describes the inward migration of Jupiter and Saturn, driven by tidal interactions with the proto-planetary disk. Saturn forms a bit later but migrates in faster, catching up to Jupiter near Mars' current orbit and trapping Jupiter in a 3/2 mean motion resonance. Locked in this resonance, they change direction and migrate back out towards their present-day orbits. This neatly explains why Jupiter did not continue to grow and settle closer to the Sun,

like the 'hot Jupiters' observed in many exoplanet systems. Furthermore, it helps explain the structure of the terrestrial planets, especially Mars' small mass and the asteroid belt's mix of two populations: the giants ate up half of Mars' accretion region, then destabilised outer-Solar-System asteroids to fall in to the current belt.

The Nice model (Tsiganis et al., 2005; Gomes et al., 2005; Morbidelli et al., 2005) continues the story from here, starting with Jupiter, Saturn, and the younger Uranus and Neptune all in resonant, circular, co-planar orbits – unlike their out-of-resonance and slightly off-circular ones today. Gravitational interactions with the remaining outer disk of planetesimals eventually destabilise the giant planets. From their initial small separations, Uranus and Neptune are rapidly scattered outward by close encounters. They disperse the outer disk and, in doing so, dynamically damp their orbits to their stable, well-separated ones today. In addition to these final orbits, the Nice model can also explain the late heavy bombardment of the inner Solar System by scattering asteroids into eccentric orbits, meaning potential collision courses with the terrestrial planets.

Both of these models have seen modifications and ongoing debates, but they paint a useful picture for the type of wider environment to have in mind while we study the specific giant impacts discussed above (e.g. Levison et al., 2011; Pierens et al., 2014; Mann, 2018). To tie these topics together, N-body simulations can follow the evolution of the ∼100 planetary embryos and many smaller planetesimals in the inner Solar System, under the gravitational influence of the giant planets, as illustrated in Fig. 1.3. In addition to tracking the accretion rate and composition of the final planets, studies like these (e.g. Raymond et al., 2009) also start to provide rough estimates for the specific properties of giant impacts that we might expect to occur.

In the even bigger picture, it is an increasingly exciting time to study planet formation because we are no longer restricted to looking at only our own solar system. The number of detected exoplanets has climbed into the thousands, revealing a great diversity of planets and planetary systems (e.g. Fressin et al., 2013). This is driving great new progress in planet formation modelling now that there exists a population distribution of observable outcomes, instead of our single home example without any constraints on its likelihood. Furthermore, by finding solar systems at different stages of their development, the initial conditions and intermediate stages are now available to study as contemporary phenomena – from proto-planetary disks and giant planet formation, to terrestrial evolution and even giant impacts.

So far, no conclusive evidence of giant impacts around other stars has been directly observed. However, telescopes are reaching the sensitivities required to detect both the immediate and long-term consequences of exoplanet collisions. For example:

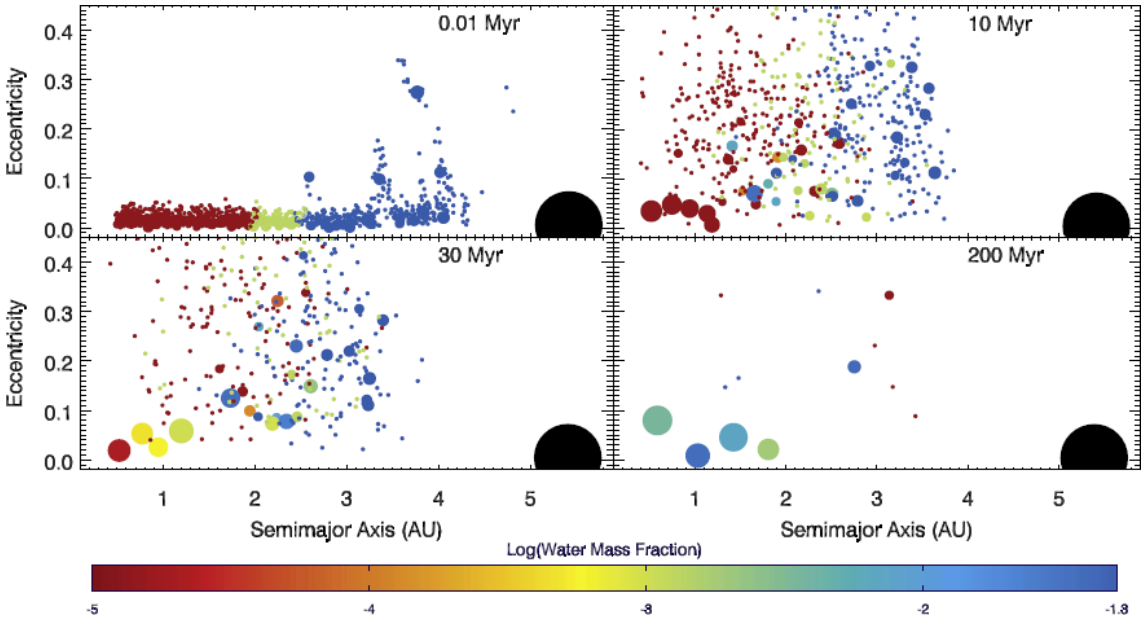


Figure 1.3: One of Raymond et al. (2009, Fig. 3)'s simulations of accretion in the inner Solar System, showing the collisional growth of the final planets from a large population of planetary embryos, and their changing composition from an initial distribution of water with heliocentric distance. The size of each body corresponds to its mass. The black circle shows Jupiter.

debris disks around a planet (Kenworthy & Mamajek, 2015), large exomoons (Kipping et al., 2014), rings of debris spread out around a star (Wyatt & Jackson, 2016), a planet's spin and obliquity (Snellen et al., 2014), and even the afterglow from an impact-heated surface (Lupu et al., 2014). In other words, most of the giant impact phenomena we see in our own solar system we can expect to begin finding elsewhere in the near future.

One consequence of particular interest – both for the evolution of exoplanets and that of the life-hosting Earth – is the fate of an atmosphere in a giant impact. The most common exoplanets around Sun-like stars have a mass between Earth's and Neptune's, and feature a remarkably diverse range of atmospheres (Fressin et al., 2013; Lopez & Fortney, 2014). The stochastic nature of giant impacts could naturally explain some of this diversity, but there is currently no consensus on the erosion caused by giant impacts (Inamdar & Schlichting, 2016). This is partly due to the lack of numerical modelling, since the standard simulations discussed in this introduction cannot resolve thin, low-density atmospheres with standard numbers of particles. For this reason, most previous studies have been restricted to using analytical or 1D approximations (e.g. Genda & Abe, 2003; Inamdar & Schlichting, 2015).

1.3 Thesis Motivations and Outline

The first goal of this thesis is to study the giant impact thought to have knocked Uranus on its side. As mentioned above, in spite of this being the earliest planetary feature attributed to a giant impact, only one previous set of 3D simulations had been run when this project began, near the very start of the field's development. With $<10^4$ SPH particles, Slattery et al. (1992)'s simulations could not resolve much beyond the average rotation of the post-impact planet. In the first half of Chapter 3, I begin reinvestigating the topic with standard-resolution simulations using 10^6 SPH particles, resolving the atmosphere, debris, and internal evolution during the impact.

However, decidedly little work has been done to test claims that features like these are 'resolved' by the simulations, and most simulation codes are not fast enough to run higher numbers of particles to check. In Chapter 2, as well as detailing the other methods used throughout the thesis, I describe development of the next-generation code SWIFT to take apt advantage of contemporary supercomputing architectures for high-resolution impact simulations. I also design an improved scheme for creating the initial conditions, which can otherwise be another computationally expensive part of a project. In the second half of Chapter 3, I use these tools to model Uranus giant impacts with an unprecedented just over 10^8 SPH particles, to test the numerical convergence of such simulations and to study the detailed erosion and angular momentum transfer of the historical collision.

Now that we can resolve orders of magnitude smaller masses in simulations, questions about atmospheric erosion are becoming feasible to answer. Chapter 4 describes an initial investigation of this virtually untouched topic, with a suite of simulations that explore a wide range of different impact scenarios onto proto-Earth targets with thin atmospheres.

Speaking of the Earth system, as well as being the subject of by far the greatest number of studies of giant impacts, the Moon also provides the best observational opportunities to test their detailed consequences. One of the Moon's primary features that an impact helps to explain is its relative lack of volatiles. Nevertheless, some volatiles are created in and escape from the lunar interior to enter its tenuous atmosphere, dominated in mass by argon. In Chapter 5, I use data from the recent LADEE mission (Benna et al., 2015) – the first to measure the lunar exosphere since Apollo 17 in 1972 – to study the newly revealed features of the system and its source inside the Moon.

Each chapter has some brief conclusions at its end, with a final summary in Chapter 6.

Chapter 2

Methods

In this chapter, I present the tools required for the giant impact simulations in this thesis. §2.1 describes the smoothed particle hydrodynamics method, followed by the SWIFT code that we developed to simulate planetary impacts while taking advantage of modern supercomputing architectures. The equations of state that model the behaviour of materials are given in §2.2. In §2.3 I describe the generation of radial planetary profiles and then present the stretched equal-area (SEA) algorithm for creating low-noise spherical arrangements of particles. The results of using SEA initial conditions are shown and compared with other schemes in §2.4. I describe the set-up of impact scenarios in §2.5, then summarise in §2.6.

2.1 Smoothed Particle Hydrodynamics

We currently lack the technology (not to mention the raw materials and funding) to test what actually happens when two planets collide in full scale experiments. This might begin to change with ever more observations of exoplanets, but for now we turn to numerical simulations. Ideally, we might create a computational model to track every atom in each planet with a particle and follow their evolution to make realistic predictions. Sadly, the daunting numbers of atoms that plague all the unhelpful analogies involving grains of sand or stars in the sky make this decidedly infeasible. This drives us to model our planets at (significantly!) lower resolution, treating simulation particles as large lumps of materials in what has become a simplified problem of fluid dynamics.

Smoothed particle hydrodynamics (SPH) is a numerical tool used in a wide range of topics in astrophysics and many other fields, from planetary impacts and supernovae to galaxy evolution and cosmology (Springel, 2010; Monaghan, 2012). As implied by the name, SPH is a Lagrangian method that models a material by following

the movement of discrete particles with time, solving the appropriate equations of hydrodynamics (and gravity). 'Smoothed' then refers to the use of a kernel to compute a hydrodynamical quantity at any point by averaging over neighbouring particles.

As a quick aside: compared with also-popular Eulerian methods that instead follow the flow of material through a grid, SPH has advantages and disadvantages, and their comparison is an entire field of study in itself (Price, 2012). Some simple benefits arise directly from the use of mass-carrying particles. First, individual particles can be tracked through time. This makes it trivial to see where any part of the final distribution of material came from and, for example, distinguish between material from an impactor and the target. Angular momentum is also precisely conserved, which is in general not the case for a grid with a fixed orientation.

Second, the resolution of SPH is inherently adaptive: in higher density regions, more particles are gathered closer together. This allows the simultaneous modelling of dense and sparse regions across a huge dynamic range. However, this can of course lead to poor resolution in very low density regions if not enough particles are used. SPH also has difficulties dealing with sharp discontinuities and the mixing of multiple materials – as is discussed in more detail below, although this can also be challenging for grid-based schemes.

Finally, when adding self-gravity to SPH, it is easy to take advantage of the extensive work employed in N-body codes to calculate gravitational forces between particles. The result is a highly robust and flexible method that is easy to adapt to new topics. Canup et al. (2013) compared SPH and CTH (a grid code) simulations of the same Moon-forming impact and found broadly similar results for the quantities of interest.

It is worth quickly noting that SPH and its viscosity implementation (described below) have been criticised by Agertz et al. (2007) as being fundamentally different from grid-based schemes and unable to resolve mixing instabilities. However, this has been vehemently refuted and the quantitative convergence of SPH has been demonstrated by several subsequent papers, most recently Tricco (2019). That is not to suggest that SPH does not still have important limitations, especially in its most basic form (e.g. Tasker et al., 2008).

The density of an SPH particle i is found by summing the masses of all particles j with masses m_j and positions $\vec{r}_j$, weighted by the kernel function, W:

$$\rho_i = \sum_j m_j \, W(r_{ij}, \, h_j) \; , \tag{2.1.1}$$

where the kernel is a function of relative separation, $r_{ij} \equiv |\vec{r}_i - \vec{r}_j|$, and the smoothing length, h.

This smoothing length determines the effective range of the kernel, and is usually allowed to vary with the local number density, such that the total mass within the kernel is the same for all particles. This is effectively equivalent to keeping a constant ratio, η, of the smoothing length to the local inter-particle separation. A Gaussian would be the natural choice for the kernel function. In practice it is better to use a function with a similar shape but that drops to zero at a finite radius, so only the closest say 50 or 100 nearest particles contribute. Here, we use the simple 3D cubic spline kernel,

$$W(w \equiv r/h) = \begin{cases} 3w^3 - 3w^2 + \frac{1}{2} & \text{if } w < \frac{1}{2} \\ -w^3 + 3w^2 - 3w + 1 & \text{if } w \geq \frac{1}{2} \end{cases} , \tag{2.1.2}$$

with 48 neighbours, corresponding to a ratio of smoothing length to inter-particle separation of $\eta = 1.2348$ (Dehnen & Aly, 2012).

Eqn. 2.1.1 is a specific application of the general SPH formula for the value of any quantity Q at any location $\vec{r}$:

$$Q(\vec{r}) = \sum_i m_i \, \frac{Q_i}{\rho_i} \, W(|\vec{r} - \vec{r}_i|, \, h_i) \, . \tag{2.1.3}$$

Each particle also carries its specific internal energy, u. The pressure and sound speed can then be computed each step following the density using the equation of state (see §2.2), $P_i(\rho_i, u_i)$ and $c_i(\rho_i, u_i)$.

The remainder of the SPH scheme can be derived directly from the density estimate, starting with the Lagrangian:

$$L = T - V = \sum_i m_i \left(\tfrac{1}{2} v_i^2 - u_i(\rho_i, \, s_i) \right) \, , \tag{2.1.4}$$

where T and V are the kinetic and potential energies, v_i is the speed, and s_i is the entropy of each particle. The detailed derivation is presented in many reviews (e.g. Price, 2012), but it is worth touching on the main ingredients here if only to appreciate the overall simplicity. In that spirit, using the Euler–Lagrange equations, $\mathrm{d}/\mathrm{d}t \, [\partial L / \partial \vec{v}] = \partial L / \partial \vec{r}$; the first law of thermodynamics (at constant entropy), $\mathrm{d}u = (P/\rho^2) \, \mathrm{d}\rho$; and the gradient of the SPH density sum, we arrive at the equations of motion and the change of internal energy for each particle:

$$\frac{\mathrm{d}\vec{v}_i}{\mathrm{d}t} = -\sum_j m_j \left(\frac{P_i}{\Omega_i \rho_i^2} \nabla_r W_i + \frac{P_j}{\Omega_j \rho_j^2} \nabla_r W_j \right) \tag{2.1.5}$$

$$\frac{\mathrm{d}u_i}{\mathrm{d}t} = \frac{P_i}{\Omega_i \rho_i^2} \sum_j m_j \, (\vec{v}_i - \vec{v}_j) \cdot \nabla_r W_j \, , \tag{2.1.6}$$

where $W_i \equiv W(r_{ij}, \, h_i)$ and $\Omega_{i,\,j}$ are the correction terms for changing smoothing

lengths.

The Euler–Lagrange derivation assumes that the Lagrangian is differentiable. This means that SPH suffers zero inherent dissipation, but also would not be able to handle discontinuities such as (physically crucial) shock waves. To solve this problem, 'artificial' viscosity terms are added, which can be explicitly translated into physically motivated forces. The standard Monaghan (1992) formulation functions as a combination of Navier–Stokes bulk and shear viscosities. This adds an extra term, Π_{ij} (defined below), to the equation of motion (2.1.5) that only applies when particles are approaching, i.e. it disappears when the velocity divergence is positive. We also include the Balsara (1995) switch, which suppresses viscosity in shearing flows, somewhat restricting its application to only the shocks where it is needed:

$$B_i = \frac{|\nabla \cdot \vec{v}_i|}{|\nabla \cdot \vec{v}_i| + |\nabla \times \vec{v}_i|} \,, \tag{2.1.7}$$

where the SPH divergence and curl of the velocity are given by

$$\nabla \cdot \vec{v}_i = \frac{1}{\rho_i} \sum_j m_j \vec{v}_{ij} \cdot \nabla_r W_i \,, \tag{2.1.8}$$

$$\nabla \times \vec{v}_i = \frac{1}{\rho_i} \sum_j m_j \vec{v}_{ij} \times \nabla_r W_i \,. \tag{2.1.9}$$

It is worth noting that more sophisticated improvements on this basic switch are an ongoing area of research, as indeed are modifications to most components of the SPH method (e.g. Cullen & Dehnen, 2010).

In practice, we first calculate the signal velocity:

$$v_{\mathrm{sig},ij} = c_i + c_j - \beta \mu_{ij} \tag{2.1.10}$$

$$\mu_{ij} = \begin{cases} \dfrac{\vec{v}_{ij} \cdot \vec{r}_{ij}}{r_{ij}} & \text{if } \vec{v}_{ij} \cdot \vec{r}_{ij} < 0 \\ 0 & \text{otherwise} \end{cases} \,, \tag{2.1.11}$$

where for planetary simulations we set the viscosity parameters $\alpha = 1.5$ (used below) and $\beta = 2\alpha$, as is typical in the literature (e.g. Reinhardt & Stadel, 2017). The added term in the equation of motion sum is then

$$\Pi_{ij} = \nu_{ij} \overline{\nabla_r W_{ij}} \tag{2.1.12}$$

$$\nu_{ij} = -\tfrac{1}{2} \frac{\alpha \, \overline{B_{ij}} \, \mu_{ij} \, v_{\mathrm{sig},ij}}{\overline{\rho_{ij}}} \,. \tag{2.1.13}$$

Finally, the timestep required by each particle is constrained using the Courant–Friedrichs–Lewy condition. This essentially requires that a sound wave must not travel further than a small fraction (C_{CFL}, here set to 0.2) of the kernel size, H_i, in

a single step:

$$\Delta t = 2C_{\mathrm{CFL}} \frac{H_i}{v_{\mathrm{sig},i}} \ . \tag{2.1.14}$$

2.1.1 Planetary SPH

Simulating planets brings up a few relevant details that we must add to this basic SPH formulation, some of which have unintended side effects. First is the inclusion of multiple materials, for example an iron core and a rocky mantle. This can be done simply by assigning each particle to be a specific material with its own equation of state (EoS), currently without any special treatment when different materials mix. i.e. the pressure and sound speed for each individual particle are functions of the density, specific internal energy, and its material type:

$$P_i = P_{\mathrm{EoS}_i}(\rho_i, u_i) \ , \tag{2.1.15}$$

$$c_i = c_{\mathrm{EoS}_i}(\rho_i, u_i) \ . \tag{2.1.16}$$

The equations of state used for the materials in this thesis are described in §2.2.

This crude treatment of multiple materials leads to undesired behaviour when a particle has neighbours of a different material. For example, near the boundary between a dense iron core and less-dense silicate mantle in a planet, a particle will be assigned a density with a kernel that sums over neighbours on both sides of the boundary. However, the particle can only use its single-material EoS to calculate the pressure, in this case causing artificially high pressures in the inner mantle and artificially low pressures in the outer core. Some attempts to mitigate these problems have been proposed, such as modifying the density sum for neighbours of different materials or implementing some check and compensation for an asymmetrically populated kernel at a vacuum boundary, but so far without a clear consensus on what constitutes a reliable improvement (e.g. Woolfson, 2007; Hosono et al., 2016; Deng et al., 2019).

Similar problems occur when the kernel is asymmetrically populated with particles. This is made obvious with the vacuum boundary at a planet's outer surface. Roughly half of the spherically symmetric kernel is empty, which artificially lowers the density and requires the smoothing length to grow in order to maintain the required number of neighbours. On the bright side, the pressure at the surface of a planet is close to zero anyway, so a further reduction should not have a strong effect on the evolution. However, it does create an artificially low density layer of outer particles, which becomes more pronounced at lower resolution.

This issue of a growing smoothing length turns into a practical problem when particles

are blasted away from the planet by a giant impact. If the debris is adequately resolved, then this does not cause any issues. However, at lower resolutions or as ejecta spread further out, a single particle can fly far away from any neighbours. As an extreme example, consider a lone particle many planetary radii away from the otherwise-intact planet. If we attempt to search for ∼50 neighbours by changing the smoothing length, then a small fractional change in the smoothing length will either fail to reach or suddenly include the millions of other particles in the planet. To avoid the computational problems this brings up, we impose a maximum allowed value for the smoothing length, corresponding to a small enough density for the pressure to be negligible.

2.1.2 SWIFT

SWIFT (SPH With Inter-dependent Fine-grained Tasking) is a hydrodynamics and gravity code for astrophysics and cosmology in open development (swift.dur.ac.uk), designed from the ground up to run fast and scale well on shared/distributed-memory architectures (Schaller et al., 2016).

For the past decade, physical limitations have kept the speed of individual processor cores constrained, so instead of getting faster, supercomputers are getting more parallel. This makes it ever more important to share the work evenly between every part of the computer so that no processors are sitting idle and wasting time.

SWIFT can function as a drop-in replacement for the Gadget-2 code, which has been widely used for cosmological and planetary impact simulations (e.g. Springel, 2005; Ćuk & Stewart, 2012), but with a >30× faster runtime on representative cosmological problems (Borrow et al., 2018). This speed is partly a result of SWIFT's task-based approach to parallelism and domain decomposition for the gravity and SPH calculations (Gonnet, 2015). By evaluating and dividing up the work instead of just the data, this provides a dynamic way to achieve good load balancing across multiple processors within a shared-memory node. The tasks are decomposed over the network in distributed memory systems using a graph-partitioning algorithm, weighting each task by the estimated computational work it requires. Combined with using asynchronous communications that are themselves treated as normal tasks, this allows the code to scale well (Schaller et al., 2016). Core routines, including the direct interactions between particles, have then been optimized using vector instructions (Willis et al., 2018).

SWIFT has a modular structure that separates different code sections for clean modifications to, for example, the physics or the hydrodynamics scheme without affecting (or even being aware of) the parallelisation and other structural components.

Any such module is switched in or out with configuration flags, allowing SWIFT to run planetary, cosmological, or any other simulation as required.

In some respects, giant impact simulations pose a harder challenge for load balancing than the cosmological simulations that SWIFT is also designed for. For a large patch of the universe, although the density becomes very much higher in a galaxy than a void, the local average density is roughly constant across a simulation box. Even a crude division of particles by position in the box to different computing cores can somewhat effectively speed up the calculation, and a more careful decomposition like SWIFT's can produce excellent strong scaling across hundreds of thousands of cores (Borrow et al., 2018).

In contrast, for a giant impact, almost all the mass is in the planet at the centre. If we use a large simulation box in order to follow the ejected debris, then the vast majority of particles can easily occupy less than 0.01% of the volume. This is similar to cosmological 'zoom-in' simulations that use a high-resolution region to focus on a single galaxy or halo. This firstly makes it harder to divide up particles between computing nodes, and secondly can require much more frequent communication. This makes it much less efficient to use a large number of cores, and difficult to fully utilise a large supercomputer to run a single planetary simulation very quickly.

Happily, most studies of giant impacts can be reframed as 'embarrassingly parallel' problems because, instead of investigating one specific collision in extreme detail, the usual aim is to study a wide range of scenarios, such as varying the impact angle and speed. For this reason, perfect scaling across many distributed-memory nodes or MPI ranks is not as important. Many impacts can each be simulated on their own single (or small number of) shared-memory node(s). For context, each particle's information requires about 200 bytes of memory, so a few tens of gigabytes as an absolute minimum will be required for the 10^8 particle simulations in this thesis, compared with the 512 GB per node on the supercomputer we used (COSMA7 at Durham University). SWIFT then uses threads and SIMD vectorisation to parallelise efficiently across the tens of cores within each node. However, as we investigate in §3.5, even for parameter-space surveys, hundreds of millions of particles may be necessary to obtain sufficiently converged results, depending on the property being studied. So improving multi-node scaling for this class of problem is a key goal for SWIFT's ongoing development.

2.2 Equations of State

A material's equation of state (EoS) describes how its pressure changes with density and thermal energy (or perhaps with the temperature, entropy, etc.). Planets, like

stars and other often-simplified objects in astronomy, are dominated by easy-to-calculate gravitational forces pulling material down, and probably-more-complicated material pressure holding it up. In the SPH simulations used here, we explicitly neglect any other forces, such as material strength, so are only concerned with gravity and pressure.

The most well known EoS is given by the ideal gas law:

$$PV = nRT \ , \tag{2.2.1}$$

which applies well to warm, low-density gases. For SPH simulations, we could use any function $P(\rho, u)$ of density and specific internal energy, no matter how complicated. Compared with the inter-particle calculations like finding neighbours or computing forces, the EoS is only used once per particle per step. So the computational cost is reliably negligible. EoS can thus be analytical equations or tabulated values for interpolation, often based on a combination of theoretical calculations and experimental results, and may include phase changes and other more complicated physics.

For the projects in this thesis, we stick to relatively simple, well-tested EoS for the relevant materials, namely: Tillotson (1962) iron and granite, for Earth-like core and mantle materials; and Hubbard & MacFarlane (1980) for Uranus' rocky core, water–ammonia–methane icy mantle, and hydrogen–helium atmosphere.

2.2.1 Tillotson Iron and Granite

This analytical EoS is probably the most widely used in impact simulations (giant and otherwise), originally developed to model hypervelocity impacts and, like many equations of state, partly motivated by nuclear weapon research.

In brief, the Tillotson EoS (Tillotson, 1962; Melosh, 2007) is based on, in compressed states, a generalisation of the Mie–Gruneisen equations and shock-wave experimental results, which include intuitive dependences on the density and energy; and extrapolations to Thomas–Fermi–Dirac equations at high pressures, which are derived from first principles in the regime where electron pressure dominates. Each material is then specified by 10 parameters, listed in Table 2.1.

In the compressed or cold state, where $\rho \geq \rho_0$ or $u < u_{\rm iv}$:

$$P_{\rm c} = \left(a + \frac{b}{\omega}\right) \rho u + A\mu + B\mu^2 \ , \tag{2.2.2}$$

where $\eta \equiv \rho/\rho_0$, $\mu \equiv \eta - 1$, and $\omega \equiv u/(u_0\eta^2) + 1$.

Material	ρ_0 (kg m^{-3})	a	b	A (GPa)	B	u_0	u_{iv} (MJ kg^{-1})	u_{cv}	α	β
Iron	7800	0.5	1.5	128	105	9.5	2.4	8.67	5	5
Granite	2680	0.5	1.3	18	18	16	3.5	18	5	5

Table 2.1: Tillotson equation of state parameters (Melosh, 2007). ρ_0 is the uncompressed density, A is effectively the bulk modulus, and u_{iv} and u_{cv} are the specific internal energies of incipient and complete vaporisation, respectively. The other parameters are less easy to interpret physically.

In the hot and expanded state, where $\rho \leq \rho_0$ and $u > u_{\mathrm{cv}}$:

$$P_{\mathrm{e}} = a\rho u + \left(\frac{b\rho u}{\omega} + A\mu e^{-\beta\nu}\right) e^{-\alpha\nu^2} , \tag{2.2.3}$$

where $\nu \equiv 1/\eta - 1$.

In the hybrid state ($\rho < \rho_0$ but $u_{\mathrm{iv}} < u < u_{\mathrm{cv}}$), the pressure is a linear combination of the two:

$$P_{\mathrm{h}} = \frac{(u - u_{\mathrm{iv}})\, P_{\mathrm{e}} + (u_{\mathrm{cv}} - u)\, P_{\mathrm{c}}}{u_{\mathrm{cv}} - u_{\mathrm{iv}}} . \tag{2.2.4}$$

At low enough temperatures and densities, the pressure becomes negative. This both is a practical problem for SPH that would pull particles to stack on top of each other, and implies that the material is no longer in a suitably fluid state but should likely be splitting into droplets. Therefore, we set the minimum pressure to 0.

The Tillotson EoS does not include an expression for the sound speed, c, but it can be derived from the partial derivative of the pressure, P, with respect to the density, ρ, at constant entropy, S:

$$c^2 = \left.\frac{\partial P}{\partial \rho}\right|_S , \tag{2.2.5}$$

which we can calculate from Tillotson's P, ρ, and specific internal energy u, using the first law of thermodynamics, $\mathrm{d}u = T\mathrm{d}S - P\mathrm{d}V = T\mathrm{d}S + (P/\rho^2)\mathrm{d}\rho$.

The sound speeds for each state come out to be

$$\begin{aligned} c_{\mathrm{c}}^2 = {} & \frac{P_{\mathrm{c}}}{\rho}\left[1 + a + \frac{b}{\omega}\right] + \frac{b(\omega - 1)}{\omega^2}\left[2u - \frac{P_{\mathrm{c}}}{\rho}\right] \\ & + \frac{1}{\rho}\left[A + B\left(\eta^2 - 1\right)\right] \end{aligned} \tag{2.2.6}$$

$$\begin{aligned} c_{\mathrm{e}}^2 = {} & \frac{P_{\mathrm{e}}}{\rho}\left[1 + a + \frac{b}{\omega}e^{-\alpha\nu^2}\right] + \left\{\frac{b\rho u}{\omega^2\eta^2}\left[\frac{1}{u_0\rho}\left(2u - \frac{P_{\mathrm{e}}}{\rho}\right) + \frac{2\alpha\nu\omega}{\rho_0}\right]\right. \\ & \left. + \frac{A}{\rho_0}\left[1 + \frac{\mu}{\eta^2}\left(\beta + 2\alpha\nu - \eta\right)\right]e^{-\beta\nu}\right\} e^{-\alpha\nu^2} , \end{aligned} \tag{2.2.7}$$

and the hybrid state is the equivalent linear combination:

$$c_{\rm h}^2 = \frac{(u - u_{\rm iv})\, c_{\rm e}^2 + (u_{\rm cv} - u)\, c_{\rm c}^2}{u_{\rm cv} - u_{\rm iv}} \,. \tag{2.2.8}$$

For SWIFT, a minimum sound speed is set using the uncompressed density and bulk modulus, $c_{\rm min} = \sqrt{A/\rho_0}$.

2.2.2 Uranus Ice, Rock, and Atmosphere

For the Uranus impacts in Chapter 3, we use the bespoke EoS of Hubbard & MacFarlane (1980, hereafter HM80). Uranus is dominated by ice, which HM80 approximate as a homogeneous mix of H_2O (56.5% by mass), CH_4 (32.5%), and NH_3 (11%). Similarly, the rocky core is taken to be a composed of SiO_2 (38%), MgO (25%), FeS (25%), and FeO (12%).

From a similar mix of experimental data and theoretical extrapolation to the Tillotson EoS, they compute the zero-temperature pressures as simple polytropes:

$$P_{0,\rm ice} = \rho^{4.067} \exp\left(-3.097 - 0.228\rho - 0.0102\rho^2\right) \tag{2.2.9}$$

$$P_{0,\rm rock} = \rho^{14.563} \exp\left(-15.041 - 2.130\rho + 0.0483\rho^2\right) \,, \tag{2.2.10}$$

with the densities in g cm^{-3} and the pressure in Mbar.

The thermal component is then added separately to give the total pressure,

$$P = P_0 + C_V \rho T \,, \tag{2.2.11}$$

where C_V is the specific heat capacity at constant volume. Neglecting any temperature dependence,

$$C_V = \sum f\nu \frac{\aleph R}{m_{\rm mol}} \,, \tag{2.2.12}$$

where the sum is over each type of molecule in the mixture, f and ν are the fractional abundance of and the number of atoms in that molecule, respectively, $\aleph$ is a modifier constant (to account for the expected vibrational degrees of freedom in this regime) of 2.1 for ice and 3 for rock, R is the gas constant, and $m_{\rm mol}$ is the molar mass.

For the hydrogen–helium atmosphere, the densities become too high for the ideal gas law to be appropriate. So, HM80 extended the ideal gas EoS to higher pressures using data from Monte Carlo models of the strongly coupled liquid state. The pressure and heat capacity are evaluated with second order polynomials of density, temperature, and combined terms. We use the many fitted parameter values in Table 1 of HM80.

Unlike the Tillotson EoS, which uses the specific internal energy, these equations are expressed in terms of the temperature. However, our SPH framework uses the specific internal energy, so we must convert between the two. The 'cold curve' internal energy at zero temperature is

$$u_0(\rho) = \int_{\rho_0}^{\rho} \frac{P_0(\rho)}{\rho^2}\,\mathrm{d}\rho\ , \tag{2.2.13}$$

which is added to the thermal component to give

$$u(\rho, T) = u_0(\rho) + C_V T\ . \tag{2.2.14}$$

We then tabulate the total pressure as a function of $\log(\rho)$ and $\log(u)$ for interpolation in the SPH code.

HM80 did not provide expressions for the sound speed. So, for simplicity and knowing that the sound speed is not a particularly sensitive parameter for SPH, we treat the H–He as an ideal gas ($c = \sqrt{\gamma P/\rho}$) and use approximate bulk moduli for the other materials: $c = \sqrt{2 \times 10^{10}\ \mathrm{dyn\ cm^{-2}}/\rho}$ for the ice mix, and $c = \sqrt{2 \times 10^{11}\ \mathrm{dyn\ cm^{-2}}/\rho}$ for the rocky core, with the density here in g cm^{-3} (Matsui, 1996) for a sound speed in m s^{-1}.

2.3 Initial Conditions

Many problems in astrophysics feature spherical symmetry, such as those involving stars or planets. Before one can simulate and study these problems with a particle-based method like SPH, each initial object must first be converted into an appropriate set of particles. Two common approaches to creating arrangements of SPH particles in spheres are: (1) to use a lattice that can be distorted until it approximately matches the required shape and densities; and (2) to relax an imperfect initial state into a fully settled one with a pre-production simulation.

A third, more recent approach is to arrange the particles analytically while accounting for the spherical symmetry from the outset, by placing particles in nested spherical shells (Saff & Kuijlaars, 1997; Raskin & Owen, 2016; Reinhardt & Stadel, 2017). These methods aim to combine the minimal computation required for lattice methods with the settled and symmetric properties from relaxation simulations. We present a comparable scheme that further ensures every particle's SPH density is within 1% of the desired value. This leads to initial conditions that are quick and simple to produce, close to equilibrium, and in which every particle has a realistic density and, therefore, pressure.

Lattice-based methods are popular because they are easy to implement and, since the inter-particle separations are uniform by construction, they can accurately match a simple density profile. This can be achieved either by stretching the lattice radially or by varying the particle mass – although keeping the masses of all particles very similar is usually desirable. However, the grid-like properties of a lattice introduce unwanted anisotropies to a problem and may be unstable to perturbations (Herant, 1994; Morris, 1996; Lombardi et al., 1999).

Furthermore, a spherically symmetric object like a planet or star features important boundaries at specific radii. Both the outer surface and any differentiated internal layers require discontinuities in density and material. The particles in a lattice are dispersed at all radii, so cannot reproduce such sharp changes at these boundaries. A similarly quick and simple alternative to lattice methods is to randomly place particles following an appropriate probability distribution function, either restricted in nested shells or anywhere in the sphere. However, these methods are noisy and result in extreme variations in local particle densities.

In SPH, the density of a particle is estimated by summing the masses of typically ∼50 nearby 'neighbour' particles, weighting by a 3D Gaussian-shaped kernel that decreases the contribution of more-distant neighbours. Thus a particle that is placed too close to another will have a higher density and not be in equilibrium. The accuracy of every particle's density is important because of how 'stiff' the equation of state (EoS) can be for a material, such as the granite planetary example we test in this chapter. This means that a slightly too-dense particle will be assigned a dramatically too-high pressure by the EoS, leading to unphysical behaviour as soon as the simulation is started. In the case of a tabulated EoS, this may also cause practical problems by pushing a particle outside the parameter space covered by the tables.

An obvious improvement on these crude analytical distribution methods is to run a simulation that iterates the initial particle positions towards a more stable state. One approach is to use an inverse gravitational field to repel the particles from each other (Wang & White, 2007). A more sophisticated version of this was developed by Diehl et al. (2015) based on weighted Voronoi tessellations. Another method is to add a damping force to reduce any transient velocities as the particles are allowed to evolve under otherwise-normal gravitational and material pressure forces. In all cases, the simulation is run until a condition is met to call the system 'relaxed', such as when the particle velocities or accelerations reach some small value.

These methods can generate particle configurations that are stable and relaxed, but at a cost of performing extra simulations. Especially for large numbers of particles, this can be a computationally expensive process and can take large amounts of time,

comparable to the final simulation for which the initial conditions are being generated. Depending on the method used, the particles may also settle to a distribution somewhat different to the desired initial profile.

The spherical symmetry and sharp radial boundaries of astrophysical objects strongly motivate the arrangement of particles in nested spherical shells. If the particles could be distributed uniformly in each shell, then no computationally expensive simulation would be required to create relaxed initial conditions. However, the equidistant distribution of points on the surface of a sphere is a challenging problem, and has been studied for applications in a wide variety of fields: from finding stable molecular structures like buckminsterfullerene to making area-integral approximations, in addition to the pure mathematical curiosity of such a trivial question in 2D (equally spaced points on a circle) becoming so complicated in higher dimensions (Saff & Kuijlaars, 1997).

Similar ideas motivated the work of Raskin & Owen (2016) and Reinhardt & Stadel (2017), who both presented algorithms for arranging particles in spherical shells. One issue with Raskin & Owen (2016)'s method is that in each shell there are a few particles with large overdensities, placing the particles slightly out of equilibrium (see §2.4). Reinhardt & Stadel (2017) divide the sphere into equal regions that can be further subdivided (using the HEALPix scheme), with the disadvantage that only sparsely distributed numbers of particles (12×4^n for $n \in \mathbb{N}$) can be placed in each shell. Furthermore, some particles in each shell show SPH densities more than 5% discrepant from the desired profile density with 10^5 particles (their Fig. 4).

Before we begin placing particles, we first detail the generation of radial profiles in §2.3.1. Then, in §2.3.2, we present an algorithm for arranging any number of particles in a spherical shell such that every particle has an SPH density within 1% of the median. Our method involves a simple division of the sphere into equal-area regions arranged in latitudinal collars, followed by slightly stretching the collars away from the poles. Concentric shells can then be set up to precisely follow an arbitrary radial density profile, taking care to align the shells with any radial boundaries. We apply this stretched equal-area (SEA) algorithm to create near-equilibrium models of planets, and present the results in §2.4.

2.3.1 Planetary Profiles

The main inputs for generating simple radial profiles are the total mass, the number of layers (e.g. core, mantle, atmosphere) and their materials, the surface pressure and temperature, and inital estimates for the outer radius and any internal boundary radii between layers, which we will refine later. To set each layer's material, we must

define: the equation of state (EoS); a conversion between temperature and internal energy e.g. the specific heat capacity and cold curve; and an expression for how heat is transferred e.g. isothermal, or adiabatic.

We start by finding the density at the surface, using the EoS with the input pressure and temperature. The change in pressure, dP, as we move inwards by a small radial distance dr across a mass dm, is given by hydrostatic equilibrium:

$$4\pi\, r^2\, dP = G\, M_{<r}\, dm \,/\, r^2 \tag{2.3.1}$$

$$dm = 4\pi\, r^2\, \rho\, dr \ , \tag{2.3.2}$$

where we have assumed a constant local density, ρ, $M_{<r}$ is the enclosed mass, and G is the standard gravitational constant. The density and temperature that provide the new pressure are then found using the EoS and the heat transfer relation. This process is repeated for the next shell until reaching the centre.

The temperature and pressure are continuous across any internal layer boundaries, so this iteration continues into the core, until the input total mass has been used up. If the input radii for the outer surface and any inner boundaries are accurate, then the final available mass will be used up just as we reach the centre. However, if any of these input radii are too large or too small, then either the mass will be used up before reaching the centre or the centre will be reached with some mass still remaining. In either case, we modify the input radii and repeat the process, until the mass discrepancy is a tiny fraction of the total mass.

For our test model of a simple Earth-mass planet in §2.4, the inputs were the Earth's mass, the Tillotson granite EoS (Tillotson, 1962; Melosh, 2007), and an isothermal temperature of 300 K, leading to an outer radius of 1.036 $R_\oplus$. We chose a constant specific heat capacity of 710 J K^{-1} kg^{-1} (Wallace et al., 1960; Waples & Waples, 2004).

The resulting density (and temperature or internal energy) profile can be the basis for creating a set of particle initial conditions. This approach is the same for more complicated planets with multiple layers and discontinuities in material and density, such as the proto-Uranus and impactor used in Chapter 3.

2.3.2 Particle Placement

The goal is to distribute a number of similar-mass particles in a sphere, such that the SPH density of every particle accurately matches a given density profile. In order to follow an arbitrary radial profile that may include sharp discontinuities, such as a core–mantle boundary or a planet's surface, it is convenient to distribute

the particles in spherical shells. The particles can then be assigned any property using other radial profiles, such as their material type and temperature or internal energy.

The two inputs for this problem are the desired total number of particles and the radial density profile. The profile is first used to find the enclosed mass at each radius. The number of particles then gives the nominal particle mass. We iterate outwards from the centre, placing particles in successive shells, following the density profile. First, we must determine the radius of each shell and how many particles are required to account for its mass (§2.3.2.1). Then, the question is how to arrange an arbitrary number of particles on a spherical shell, for which we describe our stretched equal-area (SEA) method (§2.3.2.2).

2.3.2.1 Shells and Layers

We begin by placing a tetrahedron of particles near the centre, so the first 'shell' is actually the sphere that encloses the mass of four particles. If this central sphere has radius dr_c and density ρ_c, then the thickness, dr, of a subsequent shell with density ρ is

$$dr = dr_\mathrm{c} \left(\frac{\rho_\mathrm{c}}{\rho} \right)^{1/3} . \tag{2.3.3}$$

The number of particles in a shell is then simply the mass of that shell divided by the nominal particle mass. This must be rounded to an integer, giving an actual particle mass in each shell that may be slightly different to the nominal mass. This amounts to maximum deviations of ∼1% for 10^6 total particles and ∼0.1% for 10^8. The shell thickness could be tweaked instead to enforce strictly equal particle masses. The particles in the shell are then all assigned the same properties (e.g. temperature), set by the mass-weighted mean of the profile values across the shell.

It is important to note that this shell spacing will, in general, lead to shell boundaries that do not line up with any boundaries in the profile – whether simply the outer profile edge or internal boundaries separating layers inside a differentiated planet or star. In the first case of a single-layer profile, the penultimate particle shell will typically end close to the outer edge. This leaves a thin and low-mass outermost shell with only a small number of particles that both cannot adequately cover the large area and will be too close in radius to the previous shell. For interior boundaries such as between core and mantle layers, a shell will typically straddle the discontinuity. The particles in this shell then try in vain to represent some of both materials and conditions.

To avoid these problems, we slightly tweak the input particle mass to change the mass of the first core shell and hence its radius. This influences the radii of all the

shells (Eqn. 2.3.3). We iterate the input particle masses until the boundary of the outermost shell in the first (or only) layer coincides with the profile's boundary. This leads to a slightly different total number of particles as well, but ensures a proper particle representation of the final shell in this layer, as well as of the first shell of the next layer if there is one.

A similar issue and solution arises for any subsequent boundaries. To maintain a similar particle mass in all layers, we do not change the particle mass again. Instead, we tweak the number of particles in the first shell of each outer layer. This changes the mass of that shell and hence its radius, as before. By using the thickness and density of this shell instead of the central shell in Eqn. 2.3.3, this leads to appropriate changes for all the shells in this layer. We iterate over slightly different numbers of particles in the first shell until the outermost shell's boundary coincides with the profile boundary of this layer. This is repeated at the start of each layer until a particle shell boundary matches every profile boundary both internally and at the profile's edge.

One remaining decision is at what radius to place the particles within each shell. Two average radii to consider are $r_{1/2}$, half-way between the inner and outer radii of the shell, and $r_{\text{m-w}}$, the mass-weighted mean radius. For a slowly changing density profile and/or many particles that lead to thin shells, the density is roughly constant throughout the shell and $r_{\text{m-w}} > r_{1/2}$ because the mass increases with $4\pi r^2$. In the vast majority of shells, where $dr \ll r$, these two radii are approximately equal. However, at small radii near the core, placing the particles at $r_{1/2}$ results in too-high densities, and $r_{\text{m-w}}$ gives too-low densities. We found that placing the particles at $\frac{1}{2}\left(r_{1/2} + r_{\text{m-w}}\right)$ correctly matches the mean SPH density of the particles in each shell to the profile density at that radius.

2.3.2.2 Particles on a Sphere in Equal-Area Regions

For every shell, we now have a number of particles, N, to distribute on the surface of a sphere. We begin by considering the division of a (unit) sphere into equal-area regions with small diameters, following the algorithm described by Leopardi (2007) with minor modifications. The particles can then be placed in the centre of each region. This scheme is illustrated in Fig. 2.1 and a finished example with $N = 100$ is shown in Fig. 2.2.

We further impose a stretching of the regions by latitude, to improve the particle density near the poles. Finally, each shell is randomly rotated so that the particles at the poles do not line up in successive shells.

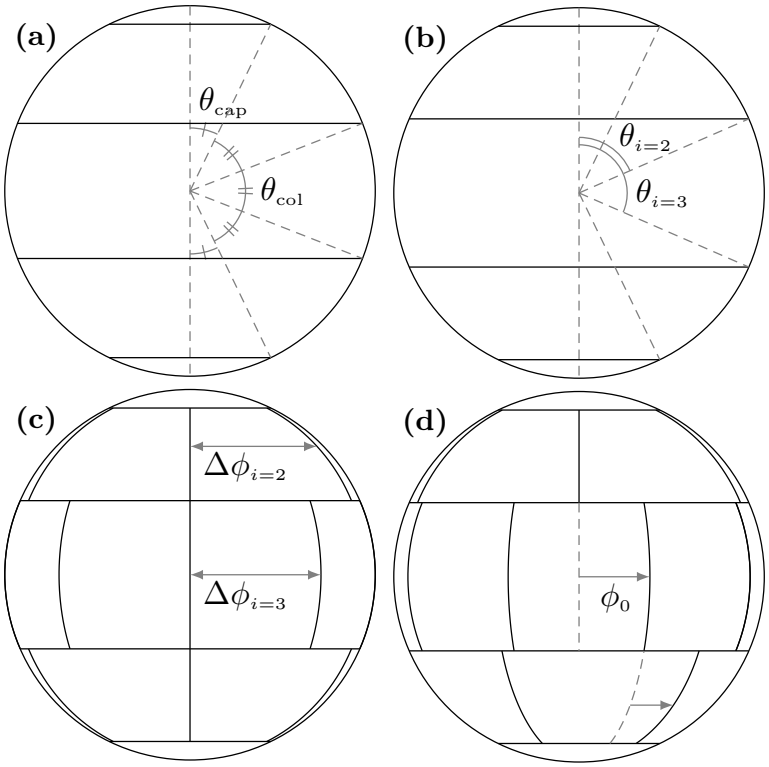


Figure 2.1: An example division of a sphere into 20 equal-area regions, demonstrating the main steps in the algorithm: **(a)** set the polar caps and the initial collar latitudes; **(b)** tweak the collar latitudes so that they each contain an integer number of regions; **(c)** divide each collar into equal regions; **(d)** rotate the collars to maximise the minimum separation of adjacent regions.

For comparison, we also test the recursive primitive refinement and parametrised spiral (RPR+PS) method described by Raskin & Owen (2016). Their method uses subdivisions of the Platonic solids for low-N shells and a spiral placement algorithm for larger numbers of particles.

For N regions on a unit sphere, the area of each one will be

$$A_{\rm reg} = 4\pi/N \ . \tag{2.3.4}$$

The bounding colatitude of a polar cap with area $A_{\rm cap}$ is

$$\theta = 2\arcsin\left(\sqrt{\frac{A_{\rm cap}}{4\pi}}\right) \ , \tag{2.3.5}$$

which for $A_{\rm cap} = A_{\rm reg}$ gives the colatitude of the single-region north pole cap, $\theta_{\rm cap}$, and south pole cap, $\pi - \theta_{\rm cap}$.

We start by dividing the rest of the sphere (between the two polar caps) into collars

with ideal initial heights of $\sqrt{A_{\text{reg}}}$. This gives the number of collars (when rounded to an integer),

$$N_{\text{col}} = \text{round}\left[\frac{\pi - 2\theta_{\text{cap}}}{\sqrt{A_{\text{reg}}}}\right] , \tag{2.3.6}$$

and the actual initial collar height (Fig. 2.1a),

$$\theta_{\text{col}} = \left(\frac{\pi - 2\theta_{\text{cap}}}{N_{\text{col}}}\right) . \tag{2.3.7}$$

We then divide each initial collar i into the closest integer number of regions. The area of each collar is

$$A_i = 4\pi\left(\sin^2\left(\frac{\theta_i}{2}\right) - \sin^2\left(\frac{\theta_{i-1}}{2}\right)\right) , \tag{2.3.8}$$

so the ideal number of regions in each collar i is

$$N_i' = \frac{A_i}{A_{\text{reg}}} . \tag{2.3.9}$$

This must be rounded to the actual integer number of regions, N_i. The cumulative discrepancy, d_i, from the ideal number of regions must be included to ensure that the total number of regions is unchanged:

$$N_i = \text{round}\left[N_i' + d_i\right] \tag{2.3.10}$$

$$d_{i+1} = d_i + N_i' - N_i . \tag{2.3.11}$$

Starting from the north pole and using the cumulative number of regions in each collar, $N_{\leq i}$, we find the final colatitude of each collar by calculating the colatitude of the cap that contains the same area as $N_{\leq i}$ regions:

$$\theta_i = 2\arcsin\left(\sqrt{\frac{N_{\leq i}A_{\text{reg}}}{4\pi}}\right) , \tag{2.3.12}$$

where $i = 1$ is the north pole cap (Fig. 2.1b).

The points in the centre of each region j in collar i then have

$$\theta = \tfrac{1}{2}\left(\theta_i + \theta_{i+1}\right) \tag{2.3.13}$$

$$\phi = \phi_0 + j\,\Delta\phi_i , \tag{2.3.14}$$

where ϕ_0 is the starting longitude and $\Delta\phi_i = \dfrac{2\pi}{N_{\leq i}}$ is the angle between adjacent points (Fig. 2.1c).

We choose the starting longitude of each collar, ϕ_0, to maximise the minimum separation between the points on adjacent collars (Fig. 2.1d). This helps to prevent

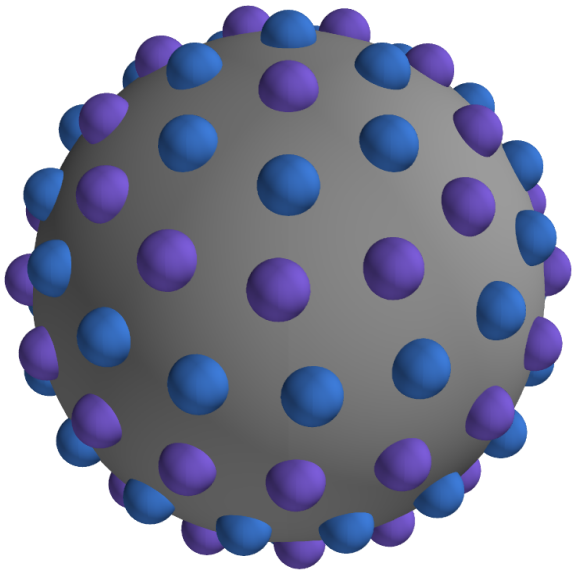


Figure 2.2: An example of 100 particles distributed on a sphere using the SEA (equal-area and subsequent latitude-stretching) method. The colours highlight each collar of particles. The SPH densities of these particles are shown by the purple points in Fig. 2.3.

local overdensities. If N_i and N_{i-1} are both odd or both even, then ϕ_0 is half the smaller of $\Delta\phi_i$ and $\Delta\phi_{i-1}$. If one is odd and the other is even, then ϕ_0 must be half of the even one's $\Delta\phi$, to prevent two particles in adjacent collars from having the same ϕ and being too close together.

Finally, we note that in successive collars with large N_i (and hence small ϕ_0) this creates a sequence of nearly adjacent particles in successive collars. To prevent this, the ϕ_0 rotation can be made with respect to a random particle in the previous collar. Thus, ϕ_0 is additionally offset by $m\,\Delta\phi_{i-1}$, where m is a random integer between 0 and N_{i-1}.

2.3.2.3 Latitude Stretching

The equal-area scheme described in §2.3.2.2 results in a small local overdensity of particles near the poles. We can make the particle density more uniform by stretching the collars near the poles. However, the collars near the equator must not be overly squashed. Therefore, the (absolute) latitude of each point, $|\frac{\pi}{2} - \theta|$, should be reduced by an amount that varies with latitude, from maximum stretching at the poles to 0 at the equator. Of course, the size of the shift at all latitudes depends on the initial size of the collars, which is set by the total number of particles. The collar height and the required shift will decrease in proportion with the square root of the number

of particles. Thus, the appropriate stretching can be given by:

$$\theta' = \theta + \left(\tfrac{\pi}{2} - \theta\right) \times aN^{-\frac{1}{2}} \exp\left[-\frac{\frac{\pi}{2} - \left|\frac{\pi}{2} - \theta\right|}{\pi\, bN^{-\frac{1}{2}}}\right] , \tag{2.3.15}$$

where $a = 0.2$ and $b = 2$ (tested for $80 \leq N \leq 10^6$). For $N < 80$, we fit a and b manually to ensure that the maximum deviation of any particle's density from the mean is less than $\pm 1\%$. This requires a to vary (non-monotonically) between 0.18 and 0.27, with b following this variation as $b = 10\,a$. These values are listed in the publicly released code, and are only relevant for the innermost one or two lowest mass shells.

2.4 Results and Discussion

In this section, we first test the arrangement of particles on an isolated spherical shell. Then, we investigate full 3D initial conditions for a simple Earth-mass planet, considering the SPH densities of the particles in their initial positions and how close they are to equilibrium when allowed to evolve.

Fig. 2.3 shows the densities of 100 particles arranged on a unit spherical shell using three different methods: Raskin & Owen (2016)'s recursive primitive refinement and parametrised spiral method (RPR+PS, specifically PS in this case) and our equal-area method without (EA) and with (SEA) the extra latitude stretching, as described in §2.3.2.2.

The RPR+PS and EA methods both show significant overdensities at the poles, with maximum deviations from the median density approaching 20% and 10% respectively. This is still a big improvement on a random distribution of particles on a shell, which leads to densities that are wrong by a factor of >10. The SEA stretching reduces the scatter to less than 1%, with typical maximum deviations of 0.5%, depending on the exact number of particles. Only 100 particles are shown here for clarity; the three methods show similar relative deviations for 10^2–10^6 particles in a single shell.

Unfortunately, this dramatic improvement of SEA over the unstretched EA method cannot be replicated for RPR+PS because the distribution of particles is not azimuthally symmetric. Stretching the RPR+PS particles at the poles reduces the overdensity for some particles but creates unavoidable underdensities for others because of their asymmetry.

To investigate how these properties of an isolated shell translate into nested shells in 3D, we now consider a full model of an Earth-mass planet with $\sim 10^5$ particles (see §2.3.1). The results from using the same three placement methods are shown in

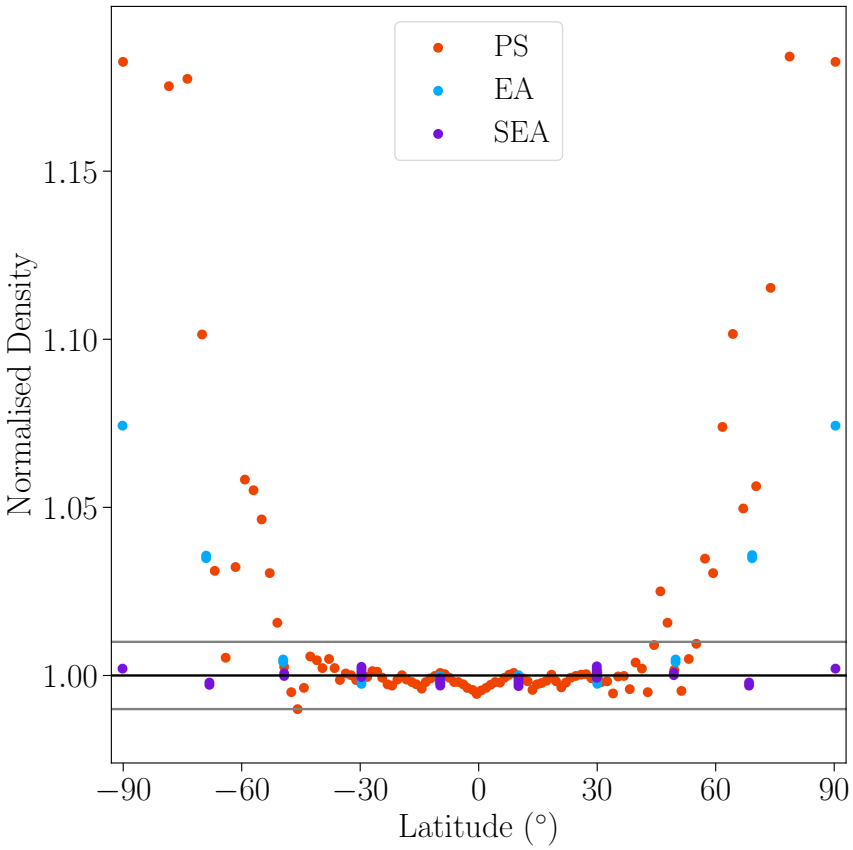


Figure 2.3: The SPH densities of 100 particles placed using the three different schemes, normalised by the median density. The particles have almost identical densities in each collar for the EA and SEA methods. The grey lines show ±1% of the median. The 3D positions of these SEA particles are illustrated in Fig. 2.2.

Fig. 2.4. As in the isolated-shell case, the RPR+PS particles show a large range of densities, with a systematic spread of particle densities more than 10% discrepant from the profile. The unstretched EA method shows similar density discrepancies around 4%, while the SEA stretching again ensures the scatter is within 1% of the profile density. These values are for a cubic spline kernel with 48 neighbours. Using another common example of the Wendland-C6 kernel with 200 neighbours yields the same qualitative results but reduces the density scatter in all cases by roughly $\frac{1}{2}$.

The underdensity of particles in the outermost shell is caused by the nature of the SPH density calculation, so is seen equally for all methods. The spherical kernel volume extends into the empty space above the planet's surface without finding any neighbours, artificially reducing the density.

It is noteworthy that the density deviations of the RPR+PS and EA methods were reduced when switching from the 2D to the 3D case, while the SEA deviations were approximately unchanged. This reflects the contributions of the particles in other

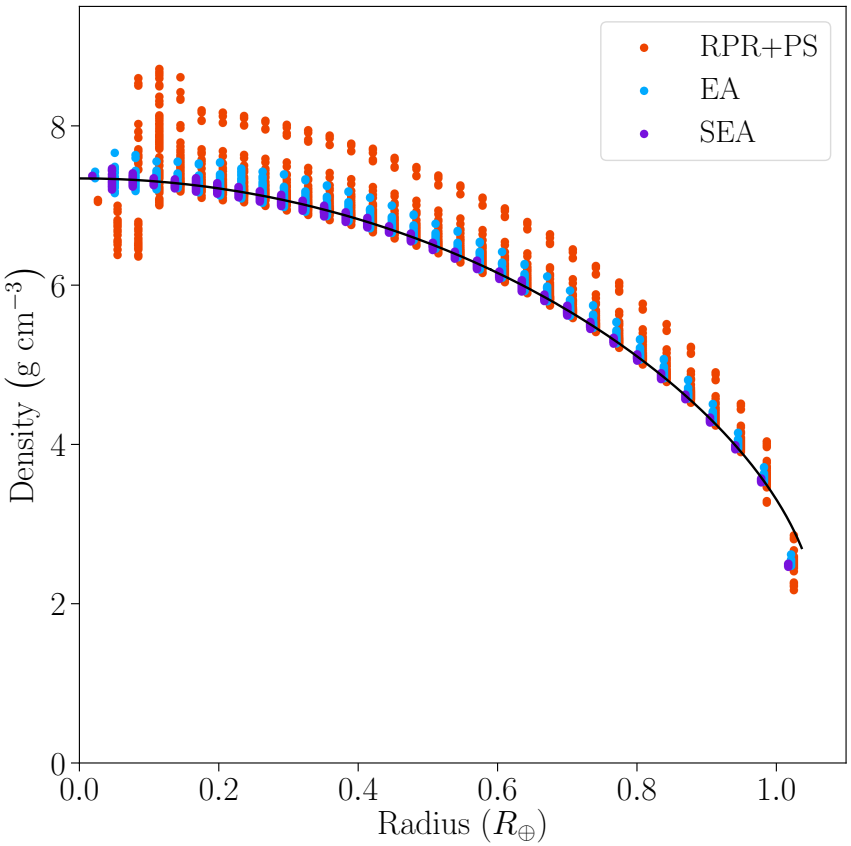


Figure 2.4: The SPH densities of $\sim 10^5$ particles placed using the three different shell schemes as labelled in the legend. The EA and RPR+PS particles are shown offset to slightly higher radii for clarity. The black line shows the input density profile, representing a simple model of an Earth-mass planet. The SEA particles' densities stay within 1% of the profile, as in the Fig. 2.3's isolated-shell case.

shells to the SPH density. The high overdensities are reduced in 3D because the nearby particles in adjacent shells are also summed over, mitigating the impact of the too-close particles in the same shell. For SEA, the particles in the randomly rotated adjacent shells are just as likely to be very slightly too close or too far as the particles in the same shell, so the density discrepancies are largely unchanged. This suggests that there would be little benefit to improving the distribution of particles within each shell beyond that of SEA, e.g. by running a relaxing simulation within each shell. Even if the particles in every isolated shell were perfectly arranged, then the imperfect contributions from adjacent-shell particles would negate any improvement. So, if even smaller density deviations were desired, then it would be necessary to consider all particles at once.

The actual success of our method is determined by how close the particles are to equilibrium when allowed to evolve in a simulation. A standard criterion for initial

conditions to be considered 'relaxed' enough for use is that the root mean square velocity, $v_{\rm rms}$, is below ∼1% of the escape speed, here $v_{\rm esc} = 11.2$ km s^{-1}. Thanks to their precise densities, the SEA particles immediately have $v_{\rm rms}$ below 0.01 $v_{\rm esc}$, and the maximum particle speed first peaks at under 0.04 $v_{\rm esc}$. ('Immediately' here meaning the fastest speeds the particles reach, soon after being allowed to evolve from a stationary start.) In comparison, a random distribution of particles in shells has initial $v_{\rm rms} = 0.2\ v_{\rm esc}$.

Most of the SEA particles' motion is caused by the previously mentioned underdensity of the outermost shell, which causes the entire planet to gently oscillate and settle into a slightly lower density profile. Because this dominates the discrepancy from an equilibrium state, the RPR+PS particles' $v_{\rm rms}$ is almost identical to SEA in spite of their comparatively noisy densities. Their maximum speed is slightly higher at 0.07 $v_{\rm esc}$. If a modified density estimator is used to fix the outer boundary problem, then a larger difference might be expected between the two methods. Planets with layers of different materials – such as the proto-Uranus and impactor in Chapter 3 – face similar SPH density problems at interior boundaries as well.

We confirmed that these relaxed SEA results are unchanged for Moon- and Pluto-mass planets (∼0.01 and 0.002 $M_\oplus$), which are less strongly gravitationally bound, making them slightly less stable. However, the pressure depends even more sensitively on density close to the low density at which the pressure is zero, for most EoS like the Tillotson (1962) used here. This exacerbates any density errors into even greater pressure discrepancies. For RPR+PS, some under-dense particles in the Pluto-mass planet are even pushed below the zero-pressure density, while the most over-dense ones get assigned a pressure over 4 times the desired value. Nevertheless, these particles can quickly be relaxed without much affecting the overall structure or $v_{\rm rms}$. SEA has the mild advantage that it avoids such issues in the first place, and requires similarly trivial computation to generate the initial conditions.

The SEAGen code for quickly generating both isolated shells and full spheres of points is publicly available at github.com/jkeger/seagen or can be installed directly with pip as the python module `seagen`.

2.5 Impact Initial Conditions

After creating our particle planets, the final step is to determine the planets' starting positions and velocities for the collision scenario we want to simulate. The relevant parameters are shown in Fig. 2.5. Depending on the situation, we may want to use different input parameters to set the initial conditions. Using conservation of energy,

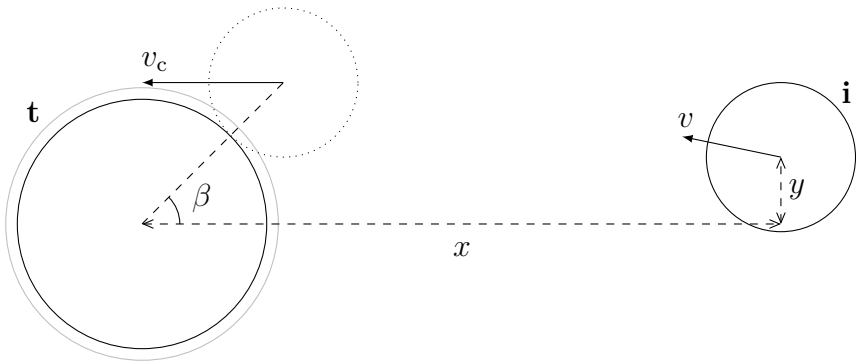


Figure 2.5: The initial conditions for an impact scenario, with the target (t) on the left with an atmosphere in grey, and the impactor (i) on the right, in the target's rest frame, rotated such that the velocity at contact is in the x direction. The simulation itself is run in the centre-of-mass and zero-momentum frame.

angular momentum, and Kepler's orbit equations, we can then derive any unknown values.

2.5.1 Contact Angle and Speed

For the atmospheric erosion simulations in Chapter 4, we care about the conditions at the moment the impactor first contacts the target. The primary inputs are thus the impact angle, β, or equivalently the impact parameter, $b \equiv \sin(\beta)$, and the speed at contact, v_c, as well as the impactor and target masses, M_i and M_t. As indicated in Fig. 2.5, these are estimated ignoring the thickness of the atmosphere and neglecting any tidal distortion before the collision.

At contact, in the target's rest frame, the distance between the body centres and the y position of the impactor are

$$r_c = R_i + R_t \tag{2.5.1}$$

$$y_c = b\, r_c \,. \tag{2.5.2}$$

For the chosen speed at contact, v_c, the velocity at infinity is

$$v_\infty = \sqrt{v_c^2 - 2\mu/r_c}\,, \tag{2.5.3}$$

where $\mu \equiv G(M_t + M_i)$ is the standard gravitational parameter. For a parabolic orbit, v_∞ is zero and $v_c = v_{esc} \equiv \sqrt{2\mu/r_c}$, the two-body escape speed. Note that for targets with atmospheres, we account for the mass of the atmosphere in M_t but ignore its thickness.

For elliptical or hyperbolic orbits, the speed, v, and y position at any earlier separation, r, can be calculated using the vis-viva equation and conservation of angular momentum, where y is always in the rotated coordinate system where v is in the $-x$ direction.

$$a = \left(\frac{2}{r_\mathrm{c}} - \frac{v_\mathrm{c}^2}{\mu} \right)^{-1} \tag{2.5.4}$$

$$v = \sqrt{\mu \left(\frac{2}{r} - \frac{1}{a} \right)} \tag{2.5.5}$$

$$y = \frac{y_\mathrm{c}\, v_\mathrm{c}}{v} \ , \tag{2.5.6}$$

where a is the semi-major axis, which is negative for hyperbolic orbits. For a parabolic orbit, $v = \sqrt{2\mu/r}$.

2.5.2 Angular Momentum

For the Uranus impacts in Chapter 3, instead of the conditions at contact, the primary inputs for each simulation are the velocity at infinity and the total angular momentum, L.

In the target's rest frame, again rotated such that the impactor is moving in the $-x$ direction, the angular momentum is

$$L = L_z = v\, y \left(M_\mathrm{i} \left(1 - m'\right)^2 + M_\mathrm{t}\, m'^2 \right) \ , \tag{2.5.7}$$

where $m' \equiv M_\mathrm{i}/(M_\mathrm{i} + M_\mathrm{t})$.

In order to allow the bodies to be distorted tidally before the impact, we set the initial separation such that, at the point on the surface of the impactor closest to the target, the gravitational force from the target planet is about 10 times smaller than that from the impactor:

$$r = \sqrt{\frac{10\, M_\mathrm{t} R_\mathrm{i}^2}{M_\mathrm{i}}} \ . \tag{2.5.8}$$

This also avoids suddenly taking the initial planets out of equilibrium by starting them right next to each other, which has been common in previous studies to avoid the extra computation during the approach.

Finally, the initial y position is derived from the input angular momentum by inverting Eqn. 2.5.7, and $x = \sqrt{r^2 - y^2}$. For simplicity, we set the initial velocity to be in the $-x$ direction, which means that the impact angle at contact will be different for each scenario.

2.5.3 Rotated Coordinates

A little extra work would allow us to rotate the coordinate system such that the velocity at contact will be in the $-x$ direction (an aesthetic choice), as depicted in Fig. 2.5. We can first find the periapsis, r_p, to get the eccentricity, e. Taking the vis-viva equation at periapsis and using Eqn. 2.5.6 to eliminate the speed gives

$$r_\mathrm{p}^2 - a\, r_\mathrm{p} + \tfrac{a v_\mathrm{c}^2\, y_\mathrm{c}^2}{2\mu} = 0$$

$$r_\mathrm{p} = \frac{a \pm \sqrt{a^2 - \frac{2 a v_\mathrm{c}^2\, y_\mathrm{c}^2}{\mu}}}{2} \tag{2.5.9}$$

$$e = 1 - r_\mathrm{p}/a\ , \tag{2.5.10}$$

which allows calculation of the true anomaly (in this case using its complement, θ) and the angle of the velocity away from the radial vector, ν:

$$\theta = \cos^{-1}\left(\frac{1 - \frac{a\left(1-e^2\right)}{r}}{e}\right) \tag{2.5.11}$$

$$\nu = \sin^{-1}\left(\frac{a^2\left(1-e^2\right)}{2ar - r^2}\right)\ . \tag{2.5.12}$$

Working through several other angles along the way, the final angle by which we want to rotate the starting x, y, and v is

$$\begin{aligned}\phi &= -(\pi - \nu - \{\pi - [\theta_\mathrm{c}\ - (\theta - \sin^{-1}\tfrac{y}{r})]\}) \\ &= \nu - \theta_\mathrm{c} + \theta - \sin^{-1}\tfrac{y}{r}\ .\end{aligned} \tag{2.5.13}$$

In the special case of a parabolic orbit, the angles can be calculated directly:

$$\theta = \pi - \cos^{-1}\left(\frac{y_\mathrm{c}^2 v_\mathrm{c}^2}{\mu r}\right) \tag{2.5.14}$$

$$\nu = \theta_\mathrm{c}/2\ , \tag{2.5.15}$$

followed by the same rotation by ϕ.

2.5.4 Time to Contact

The time taken from the initial position to contact, t_c, can be found by using the eccentric anomaly, E,

$$E_\mathrm{ell} = \cos^{-1}\left(\frac{e + \cos\theta}{1 + e\cos\theta}\right) \tag{2.5.16}$$

$$E_{\rm hyp} = \cosh^{-1}\left(\frac{e+\cos\theta}{1+e\cos\theta}\right) \tag{2.5.17}$$

$$E_{\rm par} = \tan\left(\frac{\theta}{2}\right) \tag{2.5.18}$$

and mean anomaly, M,

$$M_{\rm ell} = E - e\sin(E) \tag{2.5.19}$$

$$M_{\rm hyp} = -E + e\sinh(E) \tag{2.5.20}$$

$$M_{\rm par} = E + E^3/3 \tag{2.5.21}$$

to find the time since periapsis, $t_{\rm p}$,

$$t_{\rm p,ell,hyp} = \sqrt{\frac{|a|^3}{\mu}}\,M \tag{2.5.22}$$

$$t_{\rm p,par} = \sqrt{\frac{2\,r_{\rm p}^3}{\mu}}\,M\ . \tag{2.5.23}$$

Then $t_{\rm c} = t_{\rm p}(\theta) - t_{\rm p}(\theta_{\rm c})$.

For radial orbits, these angular equations cannot apply. Instead, the time until the point masses would contact, $t'_{\rm p}$, is

$$t'_{\rm p,par} = \sqrt{\frac{2r^3}{9\mu}} \tag{2.5.24}$$

$$t'_{\rm p,ell} = \frac{\sin^{-1}\left(\sqrt{wr}\right) - \sqrt{wr(1-wr)}}{\sqrt{2\mu w^3}} \tag{2.5.25}$$

$$t'_{\rm p,hyp} = \frac{\sqrt{(|w|r)^2 + |w|r} - \ln\left(\sqrt{|w|r} + \sqrt{1+|w|r}\,\right)}{\sqrt{2\mu|w|^3}}\ , \tag{2.5.26}$$

where w is the standard constant

$$w \equiv \frac{1}{r} - \frac{v^2}{2\mu}\ , \tag{2.5.27}$$

and we can extract $t_{\rm c} = t'_{\rm p}(\theta) - t'_{\rm p}(\theta_{\rm c})$ as before.

For the atmospheric erosion simulations in Chapter 4, we set the time until contact to be 1 hour and derive the initial separation instead of the reverse. This is more convenient given the wide variety of impact speeds we consider, compared with the Uranus impacts. Unfortunately, the equations for the mean anomaly in terms of the eccentric anomaly do not have analytical inversions. So, for practical simplicity, we iterate the input initial separation until we obtain the desired $t_{\rm c}$.

2.6 Conclusions

In this chapter, I have presented a simple method for creating spherical arrangements of particles with precise densities, and the SWIFT code for hydrodynamical simulations. These open-source tools enable the modelling of giant impacts at unprecedented numerical resolution later in this thesis, and will hopefully be of use to the wider community.

I also described the smoothed particle hydrodynamics method, the equations of state (EoS) we use for planetary materials, the creation of planetary profiles, and the set-up of initial conditions for impact scenarios.

The SEA algorithm allows the quick creation of near-equilibrium, spherically symmetric initial conditions of particles (github.com/jkeger/seagen). It ensures that every particle has an SPH density within 1% of the desired value, unlike the otherwise-similarly successful methods of Raskin & Owen (2016) and Reinhardt & Stadel (2017). This mitigates the need for expensive computation that is otherwise required to produce initial conditions that are relaxed and ready for a simulation.

The open-source SWIFT code is designed to take advantage of contemporary shared/distributed-memory architectures (swift.dur.ac.uk). For planetary giant impact simulations, this has enabled a 100–1000× improvement in the number of particles that can be used, allowing the study of brand new topics that were out of reach for lower resolution simulations.

However, while higher resolution is important, the SPH scheme used for the impact simulations in this thesis has some clear limitations and opportunities for future improvement, in addition to the well-understood density discrepancies at sharp boundaries. Material strength is neglected, hopefully justified by the dominant gravity and high temperatures involved; as is thermal conductivity and radiation, which should be insignificant over the short timescales we are modelling (Reinhardt & Stadel, 2017). The EoS are also quite simple. While the planetary profiles and initial conditions we create should be broadly realistic, as the impact progresses, some material can be shocked into densities and temperatures that should correspond to vaporisation or fragmentation, which will not be modelled correctly (Kraus et al., 2015). Now that we have a flexible code with a big improvement in computational power, we can try implementations of these missing pieces to test their practical importance in the regime of giant impacts and continue to build more realistic models in the future.

Chapter 3

Uranus Giant Impacts and Convergence

AVING prepared the tools in Chapter 2 for modelling and colliding planets using smoothed particle hydrodynamics (SPH), we first put them to use to study one of the most striking examples of the consequences of a giant impact in our solar system: Uranus' uniquely high obliquity. In the first half of this chapter, after introducing the background for this specific topic in §3.1, I describe the initial suite of impact simulations in §3.2. I present the results in §3.3 and their implications for the evolution of the post-impact planet, and summarise in §3.4.

I then turn in §3.5 to study numerical convergence with much higher resolution simulations. In §3.6 I present the results of using SWIFT and the SEA particle placement method to repeat simulations of Uranus giant impacts with up to an unprecedented 10^8 SPH particles. The conclusions are summarised in §3.7.

3.1 Knocking Over an Ice Giant

Uranus spins on its side. With an obliquity of 98° and its major moons orbiting in the same tilted plane, the common explanation is that a giant impact sent the young Uranus spinning in this new direction (Safronov, 1966). This impact might also help explain other phenomena, such as the striking differences between Uranus' and Neptune's satellite systems (Morbidelli et al., 2012; Parisi et al., 2008), the remarkable lack of heat from Uranus' interior (Stevenson, 1986; Podolak & Helled, 2012; Nettelmann et al., 2016), and its highly asymmetrical and off-axis magnetic field (Ness et al., 1986). Until now, this violent event itself has been little studied since the first SPH simulations of Slattery et al. (1992).

Uranus' equatorial ring and satellite system is remarkable in several respects. It features a set of regular, prograde, major moons, a compact inner system of rings and small satellites, and a distant group of irregular moons. The inner system and

major moons are hypothesised to have formed either from a post-impact debris disk (Stevenson, 1986; Slattery et al., 1992) or from a pre-impact proto-satellite disk that was destabilised by the post-impact debris disk and rotated to become equatorial (Morbidelli et al., 2012; Canup & Ward, 2006). The more-distant irregular satellites are thought to have been captured after the impact (Parisi et al., 2008).

The interior structure of Uranus is poorly understood. Surface emission is in approximate equilibrium with solar insolation, implying that negligible heat flows out from the planet, in striking contrast with the other giant planets (Pearl et al., 1990). This might be explained by restricted interior convection, perhaps caused by the deposition of the impactor's energy into a thin shell (Stevenson, 1986; Podolak & Helled, 2012). Such a thermal boundary layer between an outer H–He-rich envelope and an inner ice-rich layer was the crucial ingredient for the evolutionary model of Uranus produced by Nettelmann et al. (2016) that was consistent with both heat flow and gravitational moment measurements.

In contrast with terrestrial planets, the magnetic field of Uranus measured by *Voyager* 2 was not dominated by the dipole component. Higher order moments contributed significantly, and the dipole itself was both offset by approximately 0.3 Uranus radii from the centre of the planet and tilted by 60° relative to Uranus' rotation axis (Ness et al., 1986). Dynamo models producing similar magnetic fields have been constructed using a layer of convecting electrically conducting ices (Stanley & Bloxham, 2004, 2006; Soderlund et al., 2013). A feature of some of these models is the presence of a stably stratified fluid layer interior to the zone where the magnetic field is generated.

As a separate source of motivation, while the ice giants Uranus and Neptune do not receive as much attention as the nearer bodies in the Solar System, they represent the closest analogues to the mini-Neptune-class exoplanets that are the most frequently discovered by Kepler (Batalha, 2014). Given the detection efficiencies, these planets are typically found on orbits with periods of the order of 100 days (Fressin et al., 2013), but have nevertheless stimulated attempts to understand the atmospheres and histories of our ice giants in order to provide context for these exoplanet observations (Fortney et al., 2013).

The first simulations of a giant impact onto a proto-Uranus, albeit in one dimension, were done specifically to investigate whether the shock from the collision would blast away Uranus' hydrogen–helium atmosphere (Korycansky et al., 1990). This gas has a much lower mass fraction and density than the inner ice and rock material, so requires high resolution to simulate. For this reason, Korycansky et al. (1990) restricted their study to a one-dimensional spherically symmetric model where the impactor mass and some of its energy was injected into the proto-Uranus core, and

the remaining energy was placed into the atmosphere. The retained atmospheric mass was found to depend sensitively upon the amount of energy deposited directly into the atmosphere, offering the possibility that the presence of Uranus' current atmosphere might constrain allowable impact scenarios.

Building on the pioneering work of Benz et al. (1986), who used SPH simulations to model the Moon-forming giant impact on the Earth, Slattery et al. (1992, hereafter S92) produced, to our knowledge, the only paper at the time this project was published with three-dimensional hydrodynamical simulations of the hypothesised impact event that befell the proto-Uranus. While the $<10^4$ particle SPH simulations of S92 did not resolve the atmosphere, they studied collisions between a 1–3 $M_\oplus$ differentiated impactor containing iron, dunite, and ice and a similarly differentiated proto-Uranus with hydrogen and helium mixed into its ice layer. For impactor masses above 1 $M_\oplus$, they found a wide range of impact scenarios that led to a sufficiently rapidly spinning planet. Most of these collisions left ice in orbit, but only the higher angular momentum ones also placed any rock or iron into orbit, as might be expected if this material is subsequently to form any of the currently observed regular moons. Uranus' satellites comprise only $\sim 10^{-4}$ of the total system mass, corresponding to just less than the mass of a single particle in S92's simulations.

In this chapter, we present new simulations of the impact with orders of magnitude better mass resolutions than those of S92, allowing the detailed modelling of, for example, Uranus' atmosphere and its fate; the deposition of the impactor's material and energy inside Uranus; the post-impact debris disk, in particular the amount, distribution, and composition of material available for satellite formation; and the testing of S92's original conclusions for the types of impacts that could have produced the present-day spin.

3.2 Methods

3.2.1 Initial Conditions

Our proto-Uranus contains a rocky core (SiO_2, MgO, FeS, and FeO), icy mantle (H_2O, NH_3, and CH_4), and atmosphere with a solar composition mix of hydrogen and helium. The equations of state for these materials were developed by Hubbard & MacFarlane (1980, hereafter HM80) and are described in §2.2.2.

We use impactor masses of $M_\mathrm{i} = 1$, 2, and 3 $M_\oplus$ and, under the assumption that little mass escapes during the impact, set the mass of the proto-Uranus to be $14.536\ M_\oplus - M_\mathrm{i}$. The proto-Uranus is differentiated into the three distinct layers

described above. The impactor is given no atmosphere, so it has only a rocky core surrounded by an icy mantle, with the ice/rock mass ratio matching that in the proto-Uranus.

To determine the amounts of rock, ice, and atmosphere in the two bodies, we first create a spherically symmetric three-layer model for the present-day Uranus, assuming hydrostatic equilibrium. The assumed outer boundary conditions are a pressure of 1 bar and a temperature of 60 K at a radius of 3.98 $R_\oplus$. We then iterate the radii of the layer boundaries until the profile contains the desired total mass (14.536 $M_\oplus$) and a reduced moment of inertia of $I/(MR^2) = 0.21$, as described in §2.3.1. The outer temperature is slightly lower than the measured value (75 K), in order that this simple model can approach the observed reduced moment of inertia of 0.22 (Podolak & Helled, 2012). We find an ice-rich body, with 2.02, 11.68, and 0.84 $M_\oplus$ in the rock, ice, and atmosphere layers, respectively, with inner boundaries at radii of 1.0 and 3.1 $R_\oplus$. There is considerable uncertainty in the composition of Uranus; this ratio of ice to rock is comparable with that in the model of Nettelmann et al. (2013), but larger than that found by HM80 and almost twice the Solar System value adopted by S92.

The density, temperature, and pressure profiles for our Uranus model as well as the three proto-Uranus and impactor pairs are shown in Fig. 3.1. Also included are the temperature–density relations, showing our isothermal rocky cores, the approximately adiabatic power-law relation for the ice mixture used by HM80, and their fitted polynomial adiabat for the atmosphere.

One simplification present in our initial conditions is the lack of compositional mixing between the different layers. For instance, S92 included H–He mixed into the icy mantle of their proto-Uranus, and the model of Hubbard & Marley (1989) had some ice mixed into the rocky core. Given the uncertainties in the current internal structure of Uranus, and the much larger uncertainties in those of the proto-Uranus and impactor, we opt for simply differentiated bodies for these initial investigations.

The impacts we consider are violent enough to dominate over any pre-existing rotation, so our proto-Uranus (and impactor) begins without any spin. This spherical symmetry also makes the generation of initial conditions much simpler, so we leave investigating the effects of pre-impact spin for a future study.

3.2.2 Impact Simulations

These simulations were run before we had finished the development of SWIFT (§2.1.2). Instead, we used a version of the parallel tree-code HOT (Warren & Salmon,

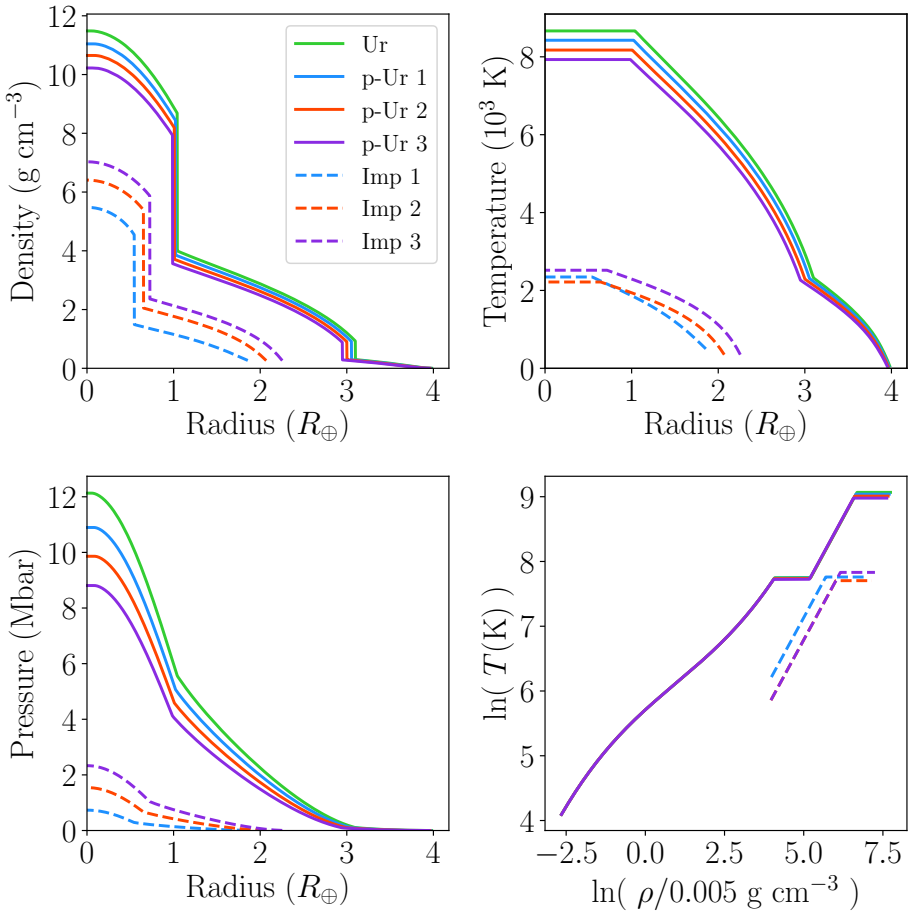


Figure 3.1: The density, temperature, and pressure profiles of our Uranus (Ur) model and the three pairs of proto-Uranus (p-Ur) and impactor (Imp) bodies. The bottom-right panel shows the temperature–density relations assumed in the various objects. The colours correspond to different masses of the impactor as labelled in the legend (in units of $M_\oplus$). The green line shows the model Uranus whose mass we split into the proto-Uranus and impactor.

1993) that has been modified to include SPH (Fryer et al., 2006) and the relevant equations of state described in §2.2. To verify the stability of our model planets given the lack of any special boundary treatments in this simple SPH formulation, we ran a simulation where the impactor misses the target but is slightly tidally disrupted, so that any problems would not be hidden in the middle of a violent impact. We confirmed that the pressure at the core–mantle boundary evolved smoothly and remained stable, showing the same 'unloading' behaviour tested by Asphaug et al. (2006, Fig. 2b).

Initial simulations of the proto-Uranus and impactor for 10,000 s in isolation were performed including a damping force to further reduce any remaining small fluc-

tuations in density. At the end of these simulations, the total kinetic energy was decreased from a fraction of $\sim 10^{-5}$ to below 10^{-6} of the total energy. This corresponds to reducing the maximum particle velocity to below 1% of the target planet's escape speed, with an average random velocity of $\sim$0.1% of the escape speed.

Prior to impact, the impactor and proto-Uranus both become distorted by the gravitational tides from the other object. The subsequent evolution can depend significantly upon these departures from sphericity at impact. Thus, for an accurate reproduction of the collision, it is necessary to start the impactor sufficiently far enough away that these tidal distortions are faithfully followed. To achieve this, we placed the impactor such that its closest particle to the proto-Uranus received a roughly 10 times larger gravitational force from the rest of the impactor than from the proto-Uranus (see §2.5.2). This amounts to separations of $\sim$22, 16, and 14 $R_\oplus$ for the 1, 2, and 3 $M_\oplus$ impactors respectively.

Separate suites of impacts were created with just over 10^5 and 10^6 particles to test the resolution dependence of our results. The angular momenta of the systems ranged from 1 to 10×10^{36} kg m^2 s^{-1}. This was achieved by changing the impact parameter while keeping the relative velocity at infinity fixed at 5 km s^{-1}, following S92 (see §2.5.2). Three head-on impacts were also simulated, one for each impactor mass. These of course cannot produce the required spin but are useful comparisons for investigating the other consequences of a collision. A set of otherwise-identical simulations with velocities at infinity ranging from 1 to 9 km s^{-1} were also performed to confirm that this choice does not significantly affect the results.

Depending on the angular momentum and impactor mass, the time taken for the impact to complete and leave a settled planet varied from roughly 1 to 7 Earth days. The simulations were stopped once the results presented in this section were not changing over timescales of 10,000 s. Using a Courant–Friedrichs–Lewy factor of 0.3 gave typical simulation timesteps of 5–10 s and 2.5–5 s for the 10^5 and 10^6 particle runs, respectively, meaning that the impact simulations typically contained $\sim 10^5$ steps.

3.3 Results

The results of the simulations are described in this section, starting with a broad description of the post-impact distribution of material. This enables us to define three mutually exclusive categories into which the particles are placed: 'planet', 'orbit' and 'unbound'. We then describe in more detail the properties of the planets that are produced, before turning our attention to the composition of the orbiting

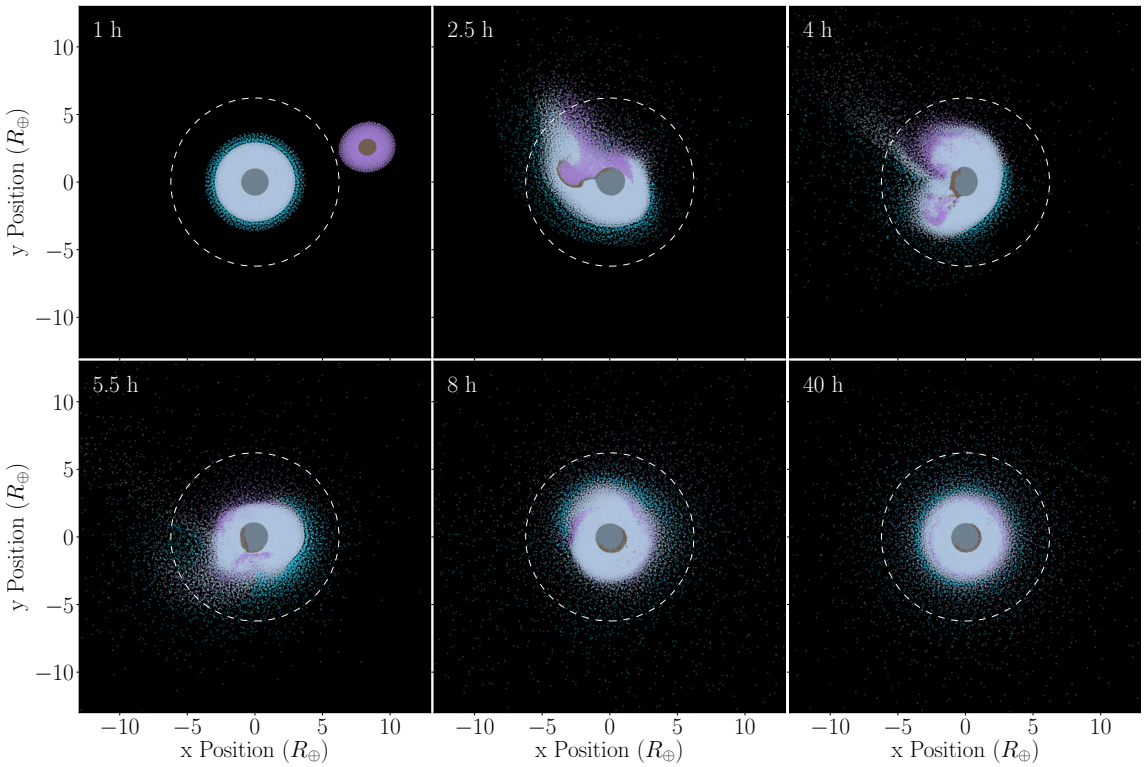


Figure 3.2: Snapshots from a low angular momentum impact simulation with a 2 $M_\oplus$ impactor and $L = 2 \times 10^{36}$ kg m^2 s^{-1}. Particles between $z = 0$ and -13 $R_\oplus$ are shown, coloured by material type and originating body. Light and dark grey show the target's ice and rock material, respectively, and purple and brown show the same for the impactor. Light blue is the target's atmosphere. The white dashed circle traces out the current Roche radius of Uranus for reference. The snapshot times are given to the nearest half hour since the start of the simulation.

debris cloud exterior to the Roche radius and the fraction of the H–He atmosphere that is retained within the Roche radius after the impact.

Given the large number of simulations, we will focus, in particular, on two 2 $M_\oplus$-impactor simulations with low ($L = 2 \times 10^{36}$ kg m^2 s^{-1}) and high ($L = 5 \times 10^{36}$ kg m^2 s^{-1}) angular momenta, as archetypal examples of ∼head-on and grazing impacts respectively. Figs. 3.2 and 3.3 show snapshots from these two giant impact simulations. These illustrate the typical features of all the impacts, with most of the impactor's rock ending up on the edge of the core of the final planet, while the impactor's ice is deposited into the outer regions of the icy mantle. At higher angular momenta, multiple passes and tidal stripping of the impactor leave more material in orbit around the final planet. Full animations of the impacts are also available at icc.dur.ac.uk/giant_impacts.

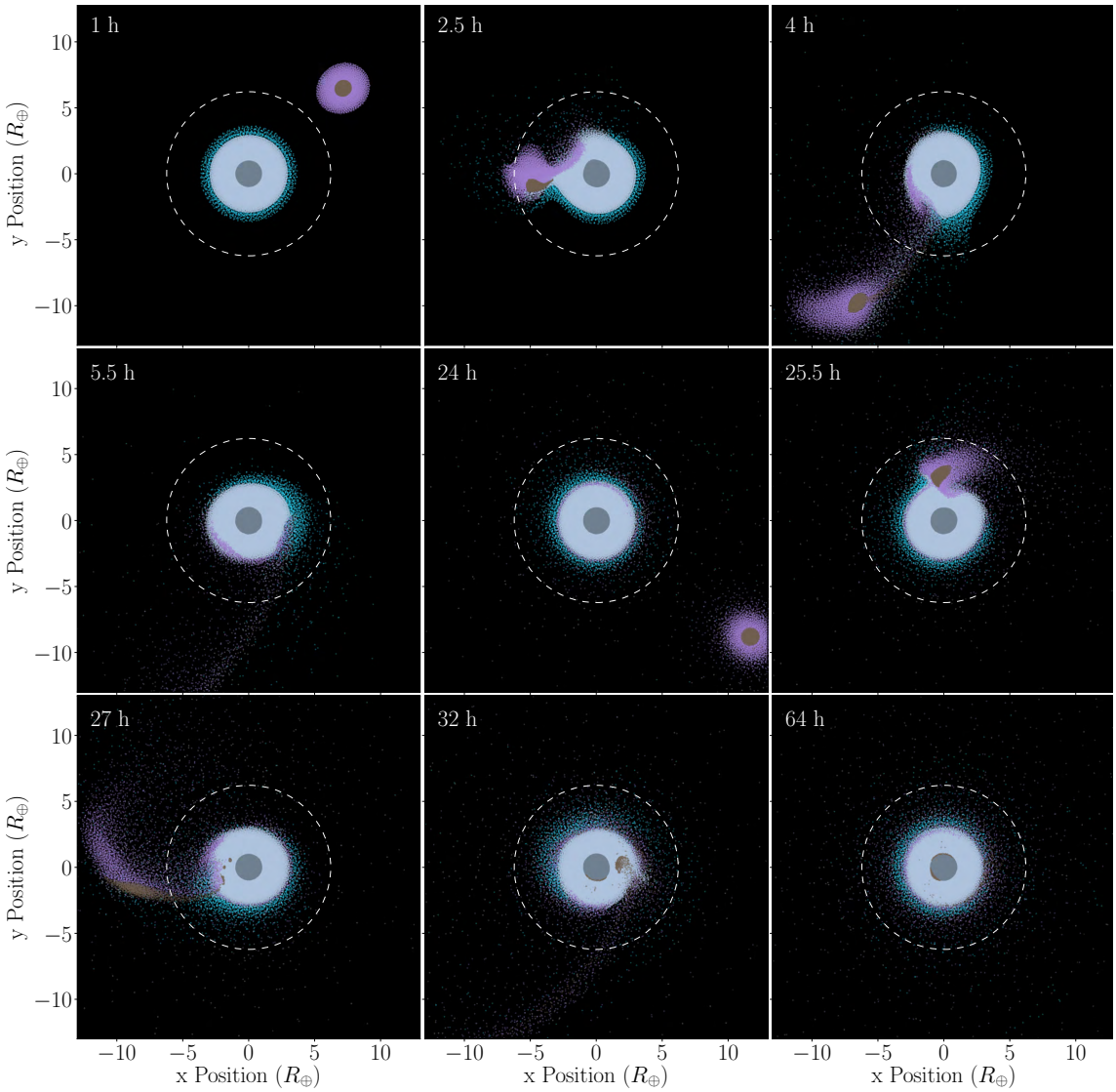


Figure 3.3: As for Fig. 3.2, but for a high angular momentum impact simulation with $L = 5 \times 10^{36}$ kg m^2 s^{-1}.

3.3.1 Material Distribution

The density profiles of the final mass distributions in the example low and high angular momentum impacts are shown in Fig. 3.4. For the more head-on collision, the impactor core is delivered more efficiently to the core of the final planet. This type of collision also places slightly more impactor ice deeper into the final planet than the relatively grazing impact. As a consequence, more of the proto-Uranus' ice and atmosphere is jettisoned into orbit around the final planet or ejected from the system entirely.

The smooth decrease in density seen for both cases in Fig. 3.4 raises the question of how to define the edge of the final planet, which is also slightly flattened due to the rotation that it has acquired. We choose to do this using a friends-of-friends

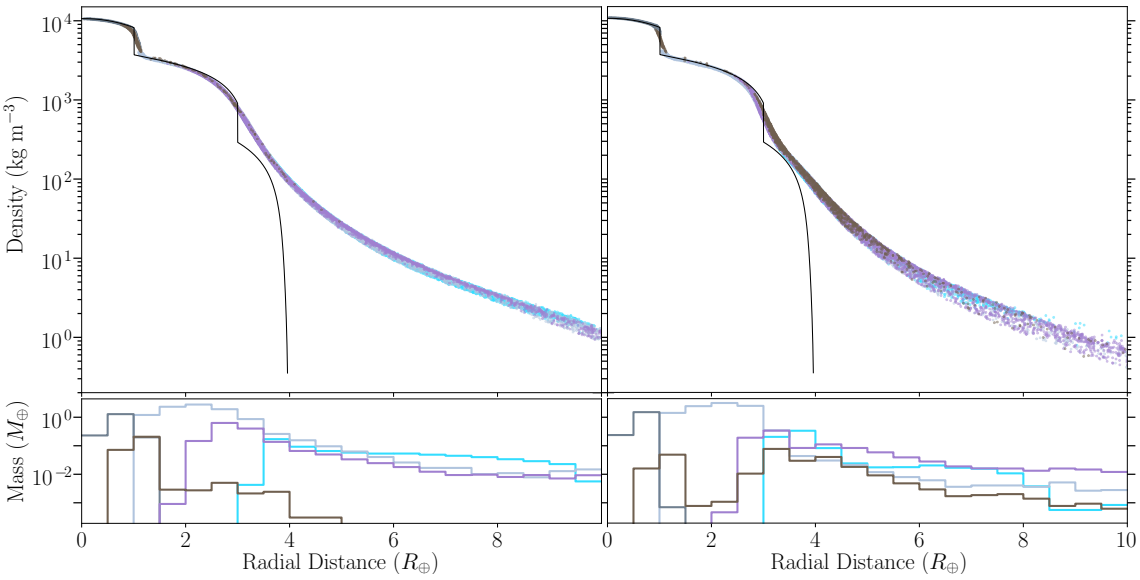


Figure 3.4: The final radial density profiles for the same relatively head-on (left) and grazing (right) impacts as in Figs. 3.2 and 3.3. The black line shows the proto-Uranus density profile. The lower panels show the mass of particles in radial bins of width 0.5 $R_\oplus$, split by material type and originating body. Light and dark grey show the target's ice and rock material, respectively, and purple and brown show the same for the impactor. Light blue is the target's atmosphere.

(FoF) algorithm (Davis et al., 1985). This links together particle pairs that are separated by less than some user-defined distance and effectively finds groups of linked particles bounded by an isodensity surface. Using a linking length of 0.3 $R_\oplus$ for the low-resolution simulations, and scaling by the inverse cube root of the particle number for the high-resolution cases leads to a final planet with a radius of $\sim$4 $R_\oplus$ and a mass that is insensitive to small changes of the linking length.

A significant amount of material external to this planet is, nevertheless, gravitationally bound to it. We will refer to this as orbiting material. The remaining mass is unbound. The orbiting material can be further divided into that within the Roche radius, which one would expect to accrete relatively quickly onto the planet, and that outside this radius, which is available to form moons. While our simulated planets have Roche radii of 5.5–5.8 $R_\oplus$ (for a satellite density of 1 g cm^{-3}), the Roche radius of present-day Uranus is 6.2 $R_\oplus$. When considering the material available for moon formation and the distribution of the post-impact H–He, we will use radii of 6 ± 0.5 $R_\oplus$ to allow for the uncertainty in the planet's mass and choice of satellite density.

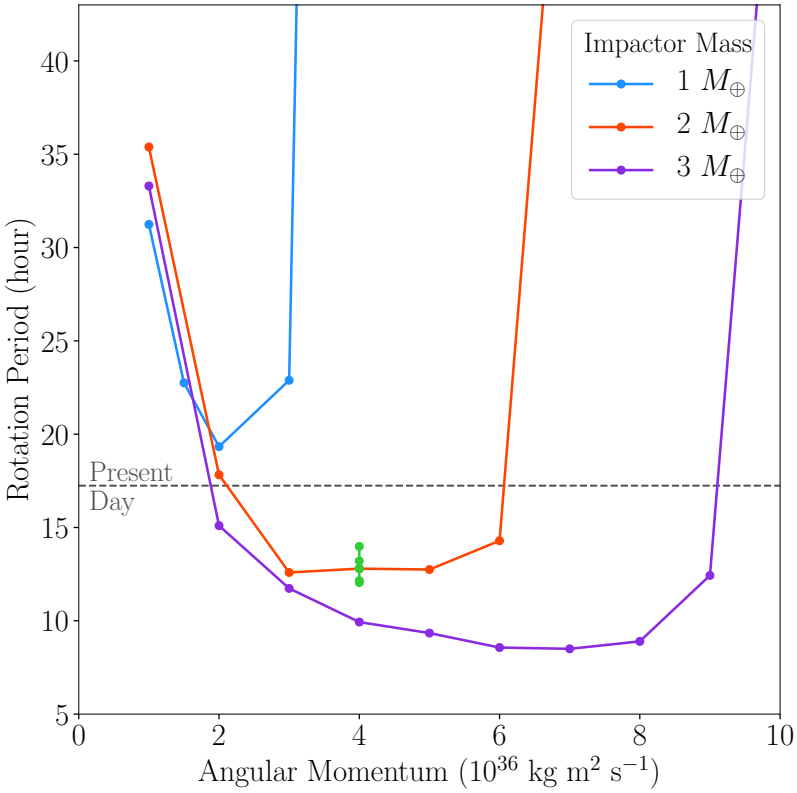


Figure 3.5: Median rotation periods for particles in the final planets produced by runs with different angular momenta and impactor masses, as given in the legend. The rotation period of each particle is calculated from its tangential velocity and distance from the z axis. All planet particles have negligible velocities in the radial and z directions. The green points show the 2 $M_\oplus$-impactor simulations with velocities at infinity of 1–9 km s^{-1} instead of the default 5 km s^{-1}. The dashed horizontal line shows the current rotation rate of Uranus of 17.24 hr (Warwick et al., 1986).

3.3.2 Resulting Planet

With the final planets defined as described in §3.3.1, we can study their rotation rates and internal structures. These properties are discussed in the following two subsections.

3.3.2.1 Rotation Rate

Fig. 3.5 shows how the rotation period varies with impactor mass and angular momentum. Despite using different proto-Uranus and impactor models from those of S92, we find broadly similar results. There is no 1 $M_\oplus$ impactor with a relative velocity at infinity of 5 km s^{-1} that can produce a sufficiently rapidly rotating planet.

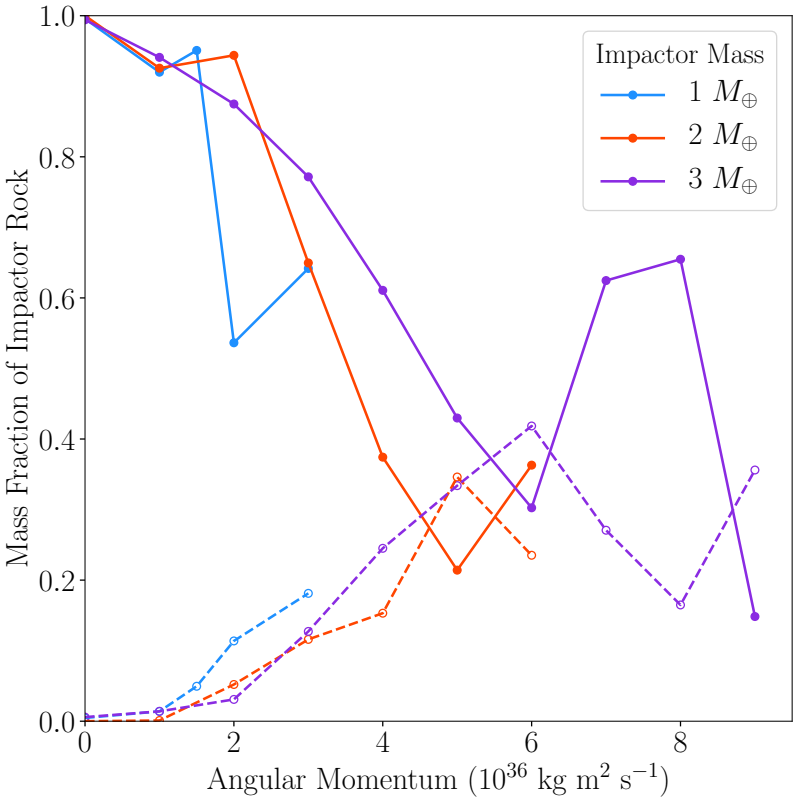


Figure 3.6: The fraction of impactor rock reaching the core of the final planet ($< 1.3\ R_\oplus$, solid lines) or deposited elsewhere in the planet (dashed lines) as a function of impactor mass (as given in the legend) and angular momentum.

Both 2 and 3 $M_\oplus$ impactors are able to satisfy this requirement, provided that the impactor is bringing an angular momentum of at least 2×10^{36} kg m^2 s^{-1}. At first, raising the angular momentum increases the final spin. However, for very high angular momentum values, to the right of the figure, the impactor starts to only graze and eventually misses the target, making it unable to transfer enough of its huge angular momentum.

Our range of simulation numbers of particles (up to $\sim 10^6$) shows that these results vary little with numerical resolution, and find them to be already well-determined with the low number of particles adopted by S92. So, the general agreement of (and any differences between) our rotation-rate results and theirs is primarily testing the different models for the colliding bodies and the materials within them, rather than showing numerical effects. This conclusion is somewhat overturned in §3.6.

3.3.2.2 Interior

The density profiles within the planet and their decomposition into material types from the two colliding bodies are shown for the low and high angular momentum

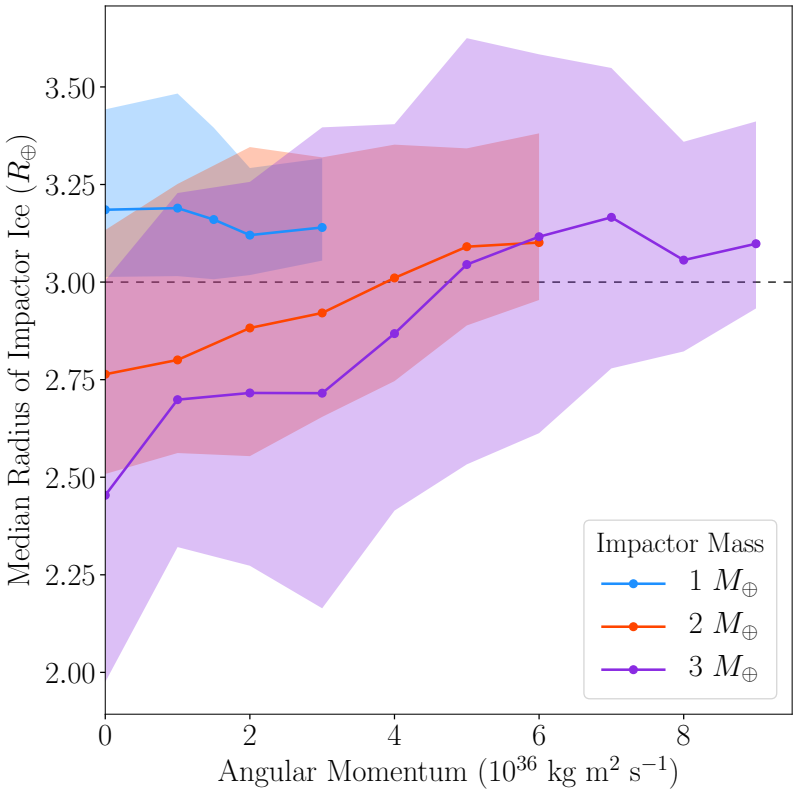


Figure 3.7: The radius of deposition of the impactor ice as a function of impactor mass and angular momentum. Shaded regions show the 1σ percentile range of the radius distributions. The dashed line shows the approximate radius of the ice-atmosphere boundary in the proto-Uranus targets.

impacts in Fig. 3.4. Considering the suite of simulations in full, Figs. 3.6 and 3.7 show the destinations of the impactor rock and ice within the planet respectively.

It is apparent from Fig. 3.6 that the head-on collisions deliver practically all of their impactor rock to the core. However, as the angular momentum is raised, the fraction of the rock in the impactor that is deposited higher up in the ice layer of the final planet or even into orbit increases significantly. The non-monotonic behaviour at high angular momenta is a consequence of an initially grazing impact sometimes leading to a much more head-on secondary collision of the core after the ice has been stripped and some angular momentum lost. Up to 40% of the rock in 2 and 3 $M_\oplus$ impactors can be left embedded in the icy mantle for sufficiently high angular momentum collisions. In our $\sim 10^6$ particle simulations, this rock is present in well-resolved, mostly spherical lumps. Such inhomogeneities will be investigated in detail with higher resolution simulations in the future, but this is beyond the scope of this initial study.

The rock that is added during the collision is generally not distributed isotropically

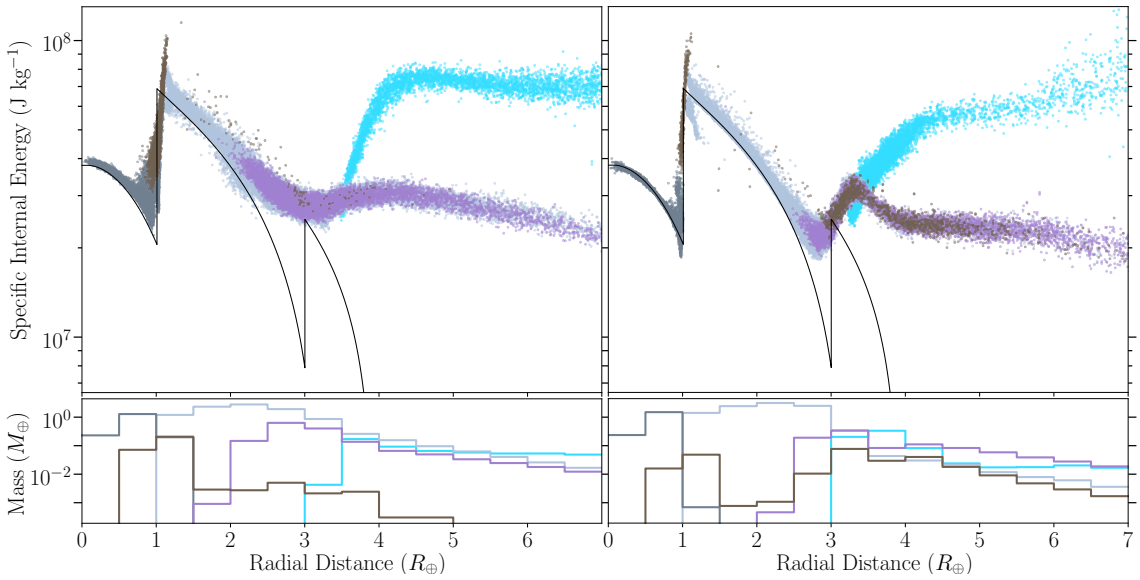


Figure 3.8: The final radial internal energy profiles for the same relatively head-on (left) and grazing (right) impacts and the same lower-panel histograms of mass per radial bin as in Fig. 3.4.

with respect to the centre of the planet. For the 2 $M_\oplus$ impactors, 90% of the delivered rock covers only ∼50% of the 4π steradians subtended at the planet's centre. This increases to ∼70% coverage for the 3 $M_\oplus$ impactors. The ice that is deposited tends to be more isotropically distributed than the rock, unless the impact is head on in which case 90% of the delivered ice subtends only ∼40%×4π steradians, independent of impactor mass.

Where this impactor ice is deposited may have profound implications for the current internal structure of and heat flow from Uranus. Fig. 3.7 shows the final destinations in radius of the impactor ice. For the 1 $M_\oplus$ impactors, the ice is mostly deposited on top of the pre-existing icy mantle, independently of angular momentum, because the impactor is not massive enough to sufficiently disturb the proto-Uranus. However, the larger projectiles are able to inject ice deeper into the final planet, particularly for the lower angular momentum collisions. These more head-on collisions also lead to a slightly thicker zone that is infiltrated by impactor ice (interquartile range spanning ∼1 $R_\oplus$) than the higher angular momentum cases, which do not penetrate as significantly into the mantle and can spread the impactor ice out into a thinner layer.

In addition to delivering mass, the impactor deposits a significant amount of energy into the final planet. The radial profiles of specific internal energy out to a little beyond the Roche radius are shown in Fig. 3.8, as well as the initial profile with its ∼adiabatic ice layer. For both low and high angular momentum collisions, the

impactor rock that reaches the edge of the final planet's core is much hotter than the largely undisturbed proto-Uranus rock. In high angular momentum collisions, a similar temperature inversion is created near the boundary between the ice and atmosphere, where the impactor ice has been delivered, creating a high-entropy layer of hot material. This sub-adiabatic energy gradient is also present in the icy mantle following low angular momentum collisions, but it is less dramatic because of the broader range of radii into which the impactor mass and energy has been deposited.

Investigating the extent and implications of this departure from adiabatic behaviour in the icy mantle compared with that required by evolution models to match the heat flow from present-day Uranus is beyond the scope of this project. However, our simulations are showing a thermal boundary layer that might suppress convection and provide a blanket to contain the heat in the central region of Uranus (Stevenson, 1986; Podolak & Helled, 2012). This layer of impactor ice could also be a compositional boundary if the icy material is not identical to that in the proto-Uranus. If these results can be usefully fed into evolution models, then this could conceivably lead to another constraint on the types of impact that are able to explain the current Uranus's thermal state and perhaps also its unusual magnetic field.

3.3.3 Orbiting Debris Field

If the moons of Uranus are to form from the debris from the collision, then it is necessary to place some rock into orbit beyond the Roche radius. Satellites would also have to form beyond the co-rotation radius of $\sim$13 $R_\oplus$ to not have their orbits decay. Using this instead of the Roche radius for our analysis reduces the amount of material available by a few tens of percent but does not change the overall conclusions. As noted by S92, this task would be made easier by having less differentiated bodies in the first place. Nevertheless, for the higher angular momentum collisions, our simulations succeed in placing significant amounts of rock and ice into the debris field. These clouds of debris are typically quite spherical rather than disk-shaped, with minimum-to-maximum axis ratios between 0.7 and 1.

The amounts of rock and ice from the impactor and the proto-Uranus in the debris cloud are shown in Fig. 3.9, as functions of impactor mass and angular momentum. This shows how the more head-on collisions send more proto-Uranus ice into orbit than impactor material. The crossover to impactor ice being more prevalent in orbit occurs at $L \approx 3 \times 10^{36}$ kg m^2 s^{-1} for impactors of mass 2 or 3 $M_\oplus$. The lowest mass impactor never manages to eject more proto-Uranus ice into orbit than impactor ice.

Grazing impacts sometimes involve multiple significant collisions or near-miss passes, creating large tidal streams of impactor material (Fig. 3.3). For a more massive

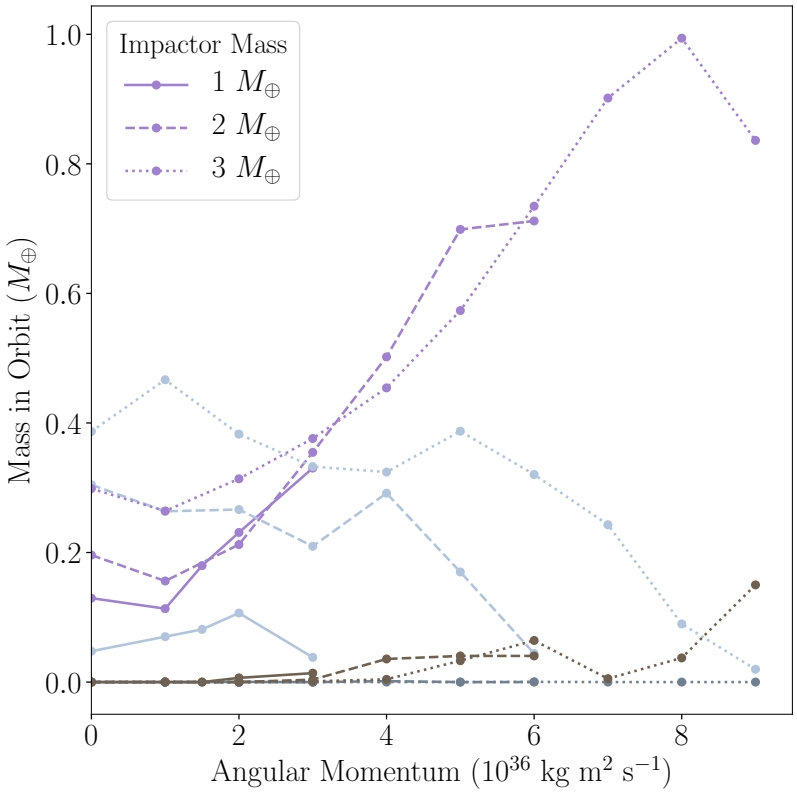


Figure 3.9: The masses of impactor and proto-Uranus material that are placed into orbit around the final planet (i.e. bound but outside a Roche radius of 6 $R_\oplus$) as functions of impactor mass and angular momentum. The line styles refer to the impactor mass and the colours to the material. Light and dark grey show the target's ice and rock material, respectively, and purple and brown show the same for the impactor.

impactor (and a correspondingly less massive proto-Uranus) the impactor's core becomes less susceptible to tidal stripping. Consequently, the higher mass impactors become less efficient at placing rock into orbit in this way. It may be that >3 $M_\oplus$ impactors would be too massive to leave any rock in orbit via this mechanism. These findings are broadly similar to those of S92; though, they were restricted to <25 rock particles in orbit and, for their more massive impactor cores, found that only <3 $M_\oplus$ impactors could be disrupted enough to leave rock in orbit.

3.3.4 Atmosphere

Most previous studies of atmospheric erosion during impacts have focused on vertical impacts onto terrestrial planets, where the atmosphere comprises a much smaller mass fraction than is present in our Uranus simulations (Ahrens, 1993). Shuvalov (2009) performed hydrodynamical simulations of oblique collisions of relatively small

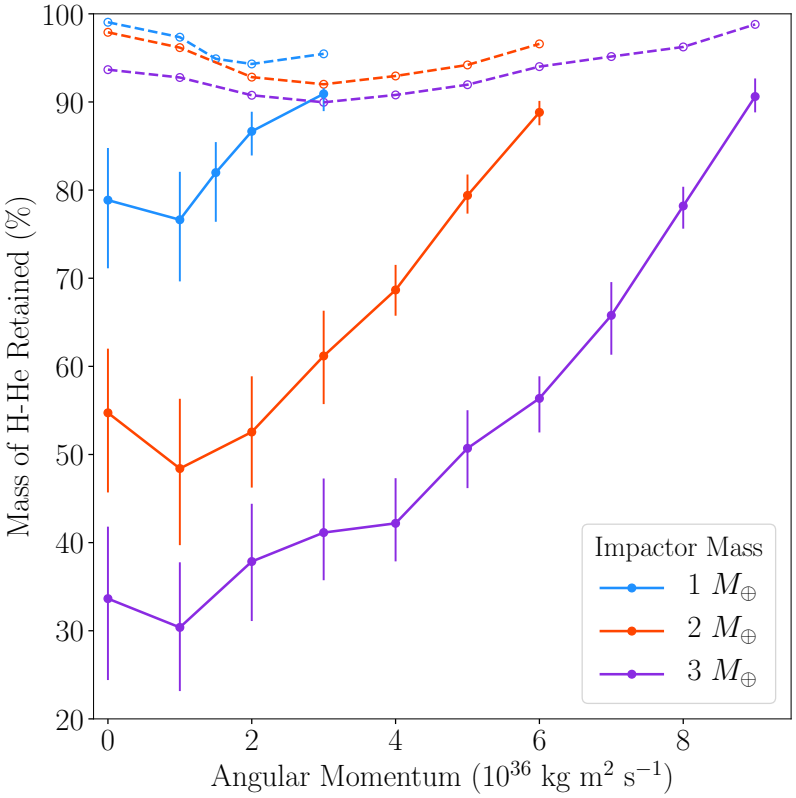


Figure 3.10: The mass fractions of the H–He atmosphere retained within a Roche radius of 6 ± 0.5 $R_\oplus$ (solid lines) and still bound to the final planet (dashed lines), as functions of impactor mass and angular momentum.

projectiles (with sizes similar to the atmosphere's height) into the Earth, finding more atmospheric erosion with more oblique impacts. For sufficiently oblique impacts, the atmospheric loss rose to all the mass above the horizon as seen from the point of impact.

For atmospheric erosion by giant impacts, Genda & Abe (2003) and Schlichting et al. (2015) used a mixture of analytical techniques and one-dimensional numerical simulations to predict that the most important factor is the speed at which the sub-atmospheric surface moves as a result of the shock wave propagating through the planet. This topic has also been little simulated in three dimensions. Liu et al. (2015) tested, to our knowledge, the only previous (at the time this project was published) three-dimensional full-planet models, with two simulations of head-on collisions on super-Earths. The simulations presented here are the first in three dimensions to quantify atmospheric erosion from giant impacts with inter-particle self-gravity as well as the first to test a range of impact angles, leaving much of this topic's huge parameter space still to be explored.

The fractions of the H–He atmosphere that are retained within the Roche radius or bound to the final planet following these giant impacts are shown in Fig. 3.10, as a

function of impactor mass and angular momentum. Most of the eroded atmosphere remains bound but can be jettisoned to large radii. There is a monotonic behaviour with larger impactors eroding more atmosphere than smaller ones, but the angular momentum dependence is more complicated. The head-on collisions retain a few more per cent of the atmosphere within the Roche radius than those with $L = 1 \times 10^{36}$ kg m^2 s^{-1}. Up to half of the atmosphere can be sent beyond the Roche radius for 2 $M_\oplus$ impactors, and this rises to 70% for $M_\mathrm{i} = 3\ M_\oplus$.

The proportion of the proto-Uranus H–He atmosphere that remains bound to the final planet is always at least ∼90%, with this minimum value being reached for intermediate values of angular momentum at $\sim 3 \times 10^{36}$ kg m^2 s^{-1}. More-grazing impacts lead to significantly higher atmospheric retention because not all the impactor's energy may be deposited at once, especially if they undergo tidal stripping and multiple less-violent collisions. As such, higher angular momentum giant impacts are less effective at eroding the atmosphere, in contrast with the trends determined by Shuvalov (2009) for the different regime of much smaller impactors. The atmosphere that is ejected by the giant impacts studied here typically originates from near to the impact site, especially in the high angular momentum cases. For the more head-on collisions, some atmosphere can also be lost on the opposite side of proto-Uranus from where the impact occurs, from the high outward velocities of the icy mantle.

3.4 Conclusions

In this first half of the chapter, we have performed SPH simulations to test the hypothesis that Uranus endured a giant impact toward the end of its formation and to investigate the consequences of such an event. We confirm the findings of S92 that the impactor needs to have a mass of greater than 1 $M_\oplus$ in order to impart sufficient angular momentum to account for Uranus' present rotation.

We also investigated where the impactor's mass and energy are deposited within the planet. Sub-adiabatic temperature gradients are typically created toward the outer regions of the icy mantle, where most of the impactor ice is deposited. Higher impact parameters can even lead to a temperature inversion near the top of the ice layer. These more-grazing collisions also leave the impactor ice further out, in a thin shell near the edge of the icy mantle, whereas ∼head-on impacts can implant significant ice up to 0.5 $R_\oplus$ further inward and less-isotropically about the centre. These findings may have important implications for understanding the current heat flow or rather the lack thereof from Uranus' interior to its surface.

With our somewhat higher resolution simulations using 10^6 particles, we see significant inhomogeneities in the deposited impactor material, and can also inspect

the composition of the debris field. The impactor's ice can be quite isotropically distributed, unlike its rocky core. While most of this rock tends to end up at the top of the core of the final planet, some small chunks become embedded within its icy mantle. For higher angular momentum impacts, significant amounts of rock and ice can be placed into orbit during tidal disruption of the impactor. The efficiency of this process is lower for 3 than 2 $M_\oplus$ impactors, since the larger impactors are more able to resist tidal stripping, but could still conceivably provide sufficient material to form Uranus' current satellites if the angular momentum of the collision exceeds 2×10^{36} kg m^2 s^{-1}.

While less than ∼10% of the H–He atmosphere of the proto-Uranus becomes unbound during the collisions, over half can be ejected to beyond the Roche radius. This atmospheric erosion occurs more in lower angular momentum collisions, where the impactor's energy is deposited all at once and some atmosphere is also lost from the antipode to the impact point.

Higher resolution simulations have allowed us to study a variety of facets of the giant impact hypothesis for producing Uranus' obliquity, including the first three-dimensional tests of atmospheric loss with inter-particle self-gravity and from off-axis giant impacts. However, even greater numbers of particles are needed to test whether these conclusions have actually converged numerically, and to resolve topics like atmospheric erosion in detail.

3.5 Convergence and High-Resolution Giant Impacts

The need to increase resolution to improve studies of existing topics was recently demonstrated by Hosono et al. (2017). Concerningly, they found giant impact simulations that gave apparently reliable results with up to 10^6 particles had not actually converged when re-tested with 10^7–10^8. Genda et al. (2015) also found incomplete convergence of disruptive impact simulations with up to 5×10^6 particles.

A numerically converged result is not necessarily physically correct. For example, several studies (e.g. Woolfson, 2007; Deng et al., 2019) have pointed out the difficulties for SPH in modelling the interaction of multiple materials or the treatment of density discontinuities, which may not be immediately fixed by higher resolutions. That said, it is crucial that we at least obtain a reliable answer to the (imperfect or not) question that we ask the computer to solve, so convergence is an important first step.

As an example with which to investigate convergence and test the simulation tools presented in Chapter 2, we consider the giant impact that likely knocked over the planet Uranus to spin on its side. In §3.3, we ran SPH simulations to study the consequences of this violent event using 10^5 and 10^6 particles – as an improvement on the $<10^4$ particles in the single previous study by Slattery et al. (1992) almost 30 years ago. Kurosaki & Inutsuka (2019) recently explored a different, complementary part of the wide parameter space with $\sim 10^5$ SPH particle simulations, varying the entropy of the proto-Uranus target to examine the effects on the angular momentum and the debris.

Now, we use the SWIFT code and the SEA particle placement method to repeat simulations of Uranus giant impacts from §3.3 using 10^5 up to 10^8 SPH particles, to test the convergence of the erosion of the atmosphere, the ejection of rocky material into the debris disk, and the post-impact planet's rotation rate.

3.6 Results

We focus on three science-motivated questions about the Uranus giant impact: (1) How much atmosphere is ejected from the system? (2) How much rocky material is placed into orbit? (3) What is the post-impact rotation period of the planet? We repeat two of the simulations from §3.6 with $\sim 10^5$, 10^6, 10^7, and 10^8 particles to investigate how these higher resolutions compare with the current standard, and to demonstrate the simulation tools described in Chapter 2.

Fig. 3.11 shows comparisons of a typical impact simulated at different resolutions, repeating the 'low angular momentum' scenario of §3.6 (Fig. 3.2). Although the overall behaviour is encouragingly similar, details like the tidal stretching of the impactor's core and the distribution of the debris clearly cannot be fully resolved by the 10^5 or 10^6 particle simulations. Fig. 3.12 highlights some the details that can be resolved with 10^8 particles for the grazing impact of the 'high angular momentum' scenario shown in Fig. 3.3.

3.6.1 Ejected Debris

In §3.3 we found that the majority of the atmosphere survives the impact, but that a small fraction can be fully ejected. Fig. 3.13 highlights the particles that will become gravitationally unbound and escape from the system. The initial collision blasts away much of the outer atmosphere and some ice, some of which will escape but most remains gravitationally bound. The 10^7 and 10^8 particle runs show that a

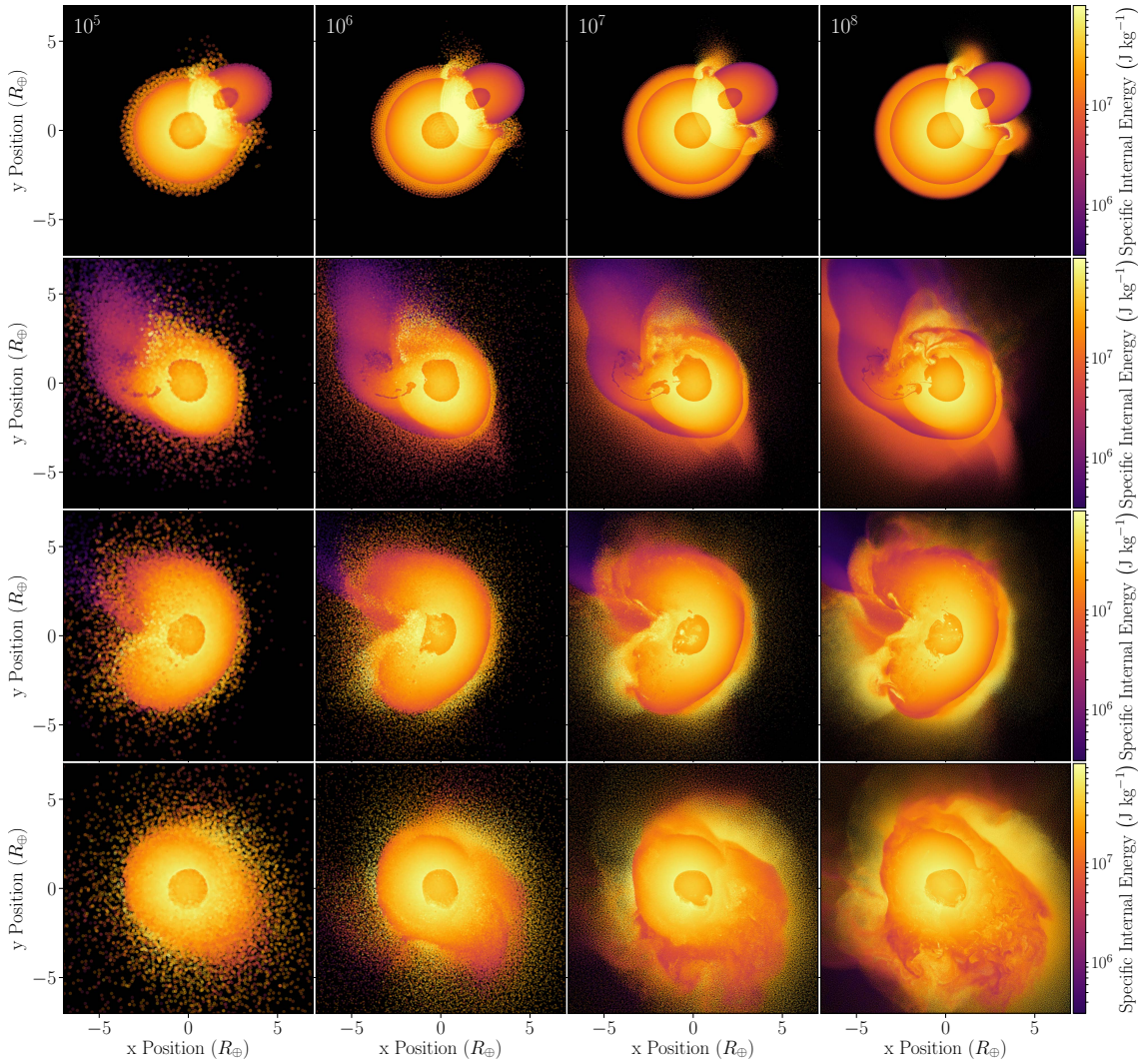


Figure 3.11: Mid-collision snapshots in the early stages of the same giant impact on Uranus at the same times from simulations with the $\sim 10^5$ SPH particles (left panels) typical in the literature, up through 10^6 and 10^7 to the 10^8 (right panels) made possible with SWIFT, resolving more of the detailed evolution of both internal structure and debris. Snapshots shown are ∼2, 3, 4, and 7 hours after the start of the simulation. An animation of the highest resolution impact is available at icc.dur.ac.uk/giant_impacts.

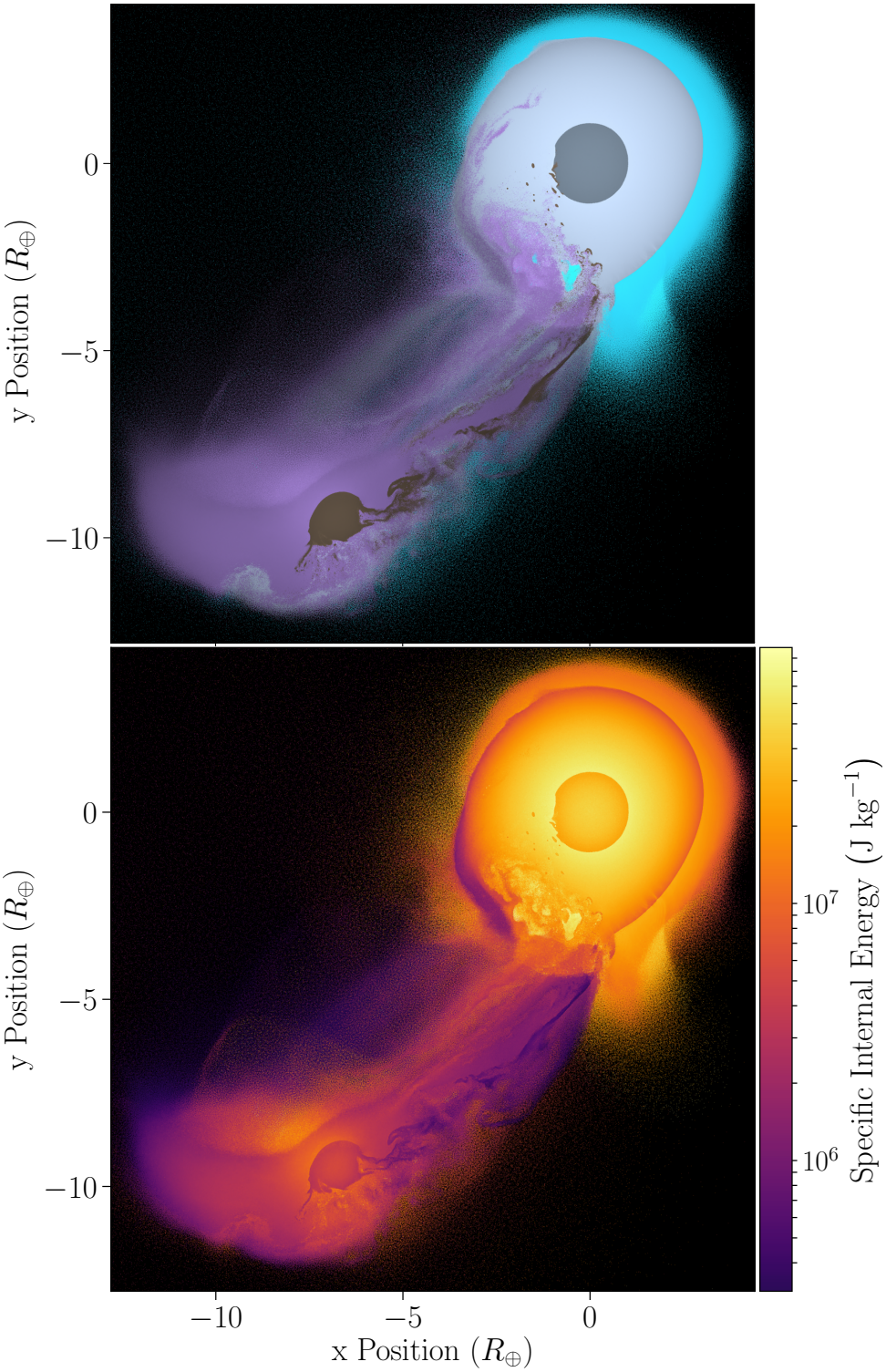


Figure 3.12: A mid-collision snapshot of a grazing impact with just over 10^8 SPH particles coloured by their material and internal energy, showing some of the detailed evolution and mixing that can now be resolved. In the top panel, light and dark grey show the target's ice and rock material, respectively, and purple and brown show the same for the impactor. Light blue is the target's H–He atmosphere.

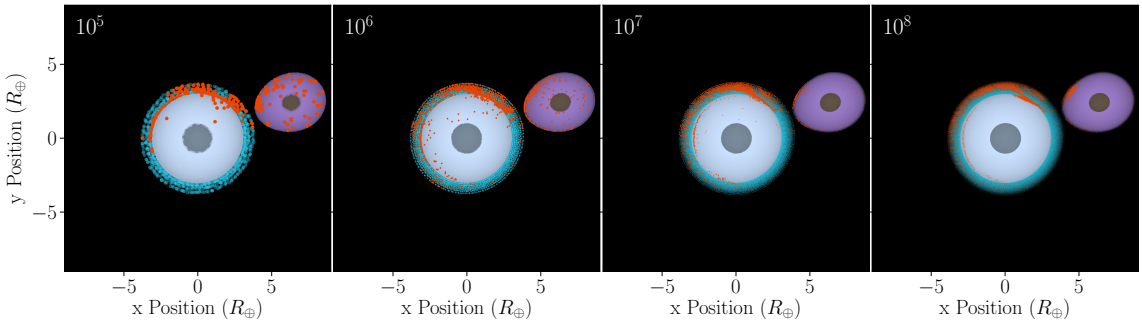


Figure 3.13: The particles that will become unbound and escape the system, highlighted in orange on a pre-impact snapshot from the same simulations with $\sim 10^5$–10^8 SPH particles as in Fig. 3.11. Only particles in a thin cross-section are shown for clarity. The colours are the same as in Fig. 3.12. The times at which these particles become unbound are shown in Fig. 3.14. The total mass lost is similar in all cases, but 10^5–10^6 particles fail to resolve the detailed results.

deeper shell of now-exposed particles then gets ejected during the subsequent violent oscillations as the impactor remnants fall back in and the planet slowly starts to settle.

The time at which this ejected material becomes unbound in each simulation is shown in Fig. 3.14. Significant mass is blasted off the planet even several hours after the initial collision in all cases. The 10^7 and 10^8 simulations closely agree that 9% of the total atmosphere mass escapes. The 10^5 and 10^6 simulations differ (non-systematically) with 8% and 12%, respectively. This suggests that atmospheric erosion has converged by $\sim 10^7$ particles in this case. On the positive side, although the lower resolution simulations do not show perfectly converged behaviour, for answering the practical question of how much atmosphere is lost, all simulations give a qualitatively similar answer of $\sim$10%.

Most studies of impact erosion use analytical or one-dimensional models to estimate the ejected atmosphere given a certain ground speed from the shock induced by the impact (e.g. Inamdar & Schlichting, 2016). In our simulations, the initial shock removes 8% of the atmosphere, then an additional 1% is lost in the subsequent sloshing. So, much like the minor resolution dependence, general conclusions about the fraction of atmosphere lost to an impact of this scale are unlikely to change. However, for more precise studies, smaller atmospheres, and perhaps other impact scenarios, this process should not be ignored.

For comparison, also shown in Fig. 3.14 is the mass of unbound ice. The 10^7 and 10^8 simulations again give similar final answers, but do not show the same behaviour at earlier times. The lower resolution simulations are discrepant by more than a factor

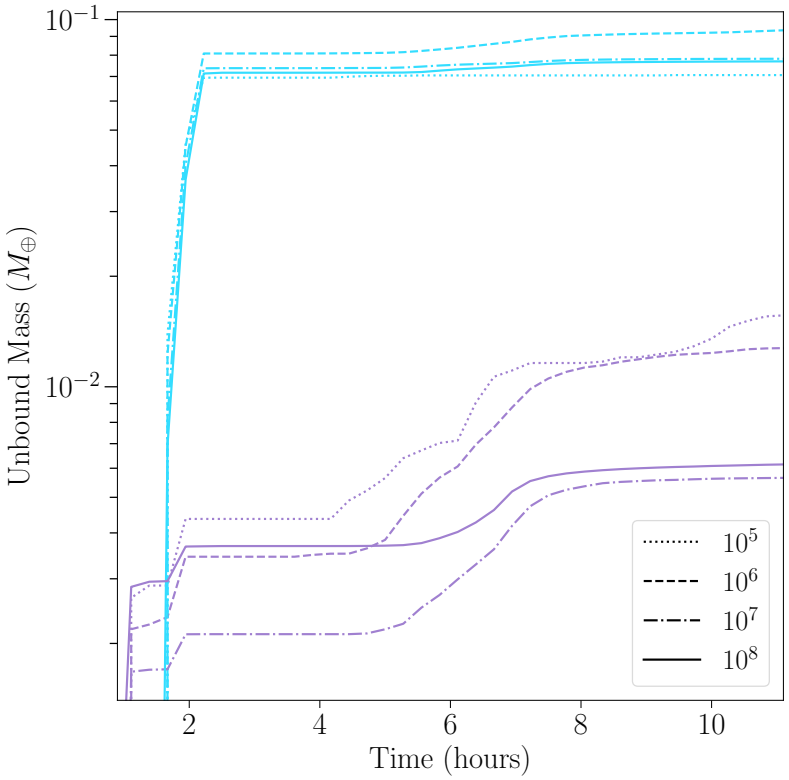


Figure 3.14: The time evolution of the mass of gravitationally unbound atmosphere (light blue) and impactor-ice (purple) material that is ejected from the system – the same particles highlighted in Fig. 3.13 – for the different resolution simulations.

of 2. It seems plausible that this quantity is approaching convergence, but without more particles than 10^8 (or checking $10^{7.5}$), it is clearly not safe to assume that this is a fully reliable result.

These quantities are summarised in Fig. 3.15 at 14 hours as a function of the number of particles, showing by how much each simulation differs from the highest resolution. That the eroded atmosphere appears closer to convergence than the ice is not surprising given the order-of-magnitude lower mass of ejected ice, meaning correspondingly fewer particles are involved in attempting to resolve the process – as can be interpreted from the sizes of the error bars.

As an example of a property that has certainly not converged, we also plot the mass of rock that is ejected into orbit in a debris disk beyond the Roche radius, where it might be available for accretion into satellites. Not only do the 10^7 and 10^8 simulations not agree, they differ by more than the lower resolutions with no semblance of convergence. The corresponding number of orbiting rock particles in each simulation is only 4, 80, 1000, and 20,000, respectively. So, especially for a messy ejection process that is widely spread out in both space and time, it is

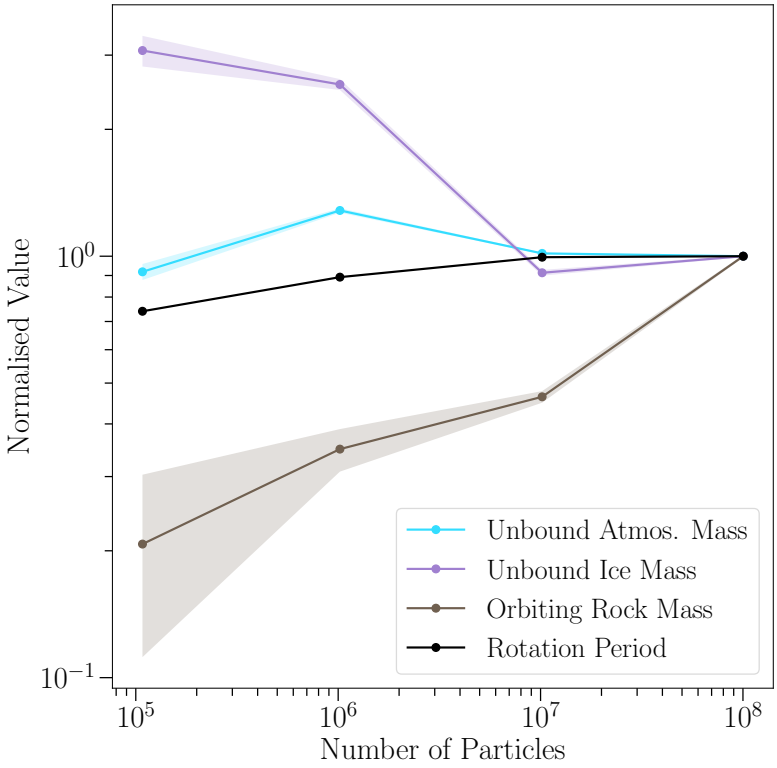


Figure 3.15: The change with resolution of the masses of unbound atmosphere and ice, the mass of rock placed into orbit beyond the Roche radius, and the planet's rotation rate, demonstrating a range of apparent (un)convergence. Each property is normalised by the highest-resolution result to show the relative differences. The shaded regions show the 1σ errors, some of which are too narrow to see. The rotation period appears to have converged by $\sim 10^7$ particles, as have – with decreasing certainty – the unbound atmosphere and ice masses, while the orbiting rock mass has not.

not surprising that 1000 or fewer particles are far from able to sufficiently resolve what happens. It is possible that the 10^8 simulation has already fully resolved and converged on this result, but our only means of checking this – running even higher resolution simulations – we leave for future studies where this is a targeted science result. In comparison, the masses of orbiting ice and atmosphere particles in the debris are much higher, and converge similarly to the unbound atmosphere mass.

3.6.2 Rotation Period

The rotation period of the post-impact planet is a large-scale bulk property involving a large majority of the particles, so one might expect it to have converged by fairly

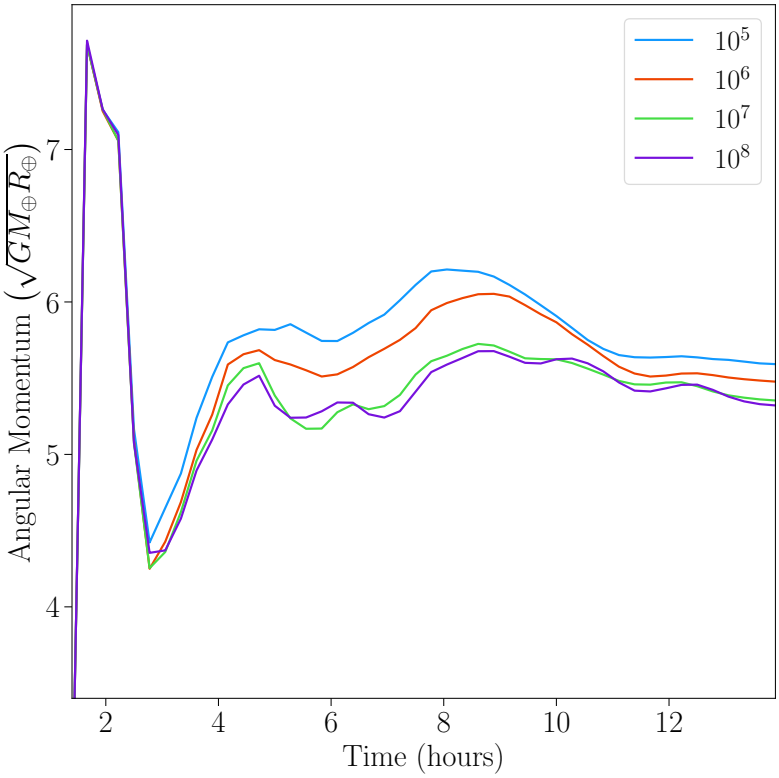


Figure 3.16: The early time evolution of the planet's angular momentum for the different resolution simulations, summed over all particles within the Roche radius of 6 $R_\oplus$. The standard- and high-resolution simulations begin to differ as the ejecta from the initial impact fall back to the planet.

low particle numbers. However, as shown in Fig. 3.15, the 10^5 and 10^6 simulations find rotation periods of 14.7 and 17.7 hours while the 10^7 and 10^8 simulations agree on a much longer period of 19.9 hours, to within 0.5% of each other. This changes our conclusions from §3.3, when it appeared that even fairly low-impact-parameter 2 $M_\oplus$ impactors could impart enough spin to explain the planet today. Assuming a similar reduction in spin for other impact scenarios at sufficiently high resolution, only a narrower range of more-grazing impacts (or more massive impactors) would be viable.

The evolution of the planet's angular momentum for each simulation is shown in Fig. 3.16, which, for simplicity, we sum over all particles within the Roche radius. The total angular momentum of the entire system remains the same in all cases, but at higher resolutions more angular momentum is transported out to the debris disk beyond the Roche radius, leaving less in the planet. All the simulations agree during the arrival and initial merging of the impactor, but their behaviour begins to diverge as the thrown-out debris (see the middle two rows of Fig. 3.11) begins to

fall back in to the planet, at around 3 hours after the start of the simulation.

Even though the total number of particles used to measure the planet's rotation rate is very large, the messy ejecta and mixing around the outer regions of the planet are significant enough to affect the overall system while also small enough to require high resolutions to model correctly. This is comparable to the effect seen by Hosono et al. (2017) where the mass of the post-impact disk did not converge as expected because of subtle differences in the detailed behaviour of re-impacting debris.

There will always be even smaller structures that are not properly resolved, but their ability to alter the rest of the system will also decrease, so appropriate-scale quantities should stay converged. However, properties such as small-scale turbulent mixing and the emergence of smaller structures may never converge without the addition of regularising physics such as diffusion or viscosity mechanisms (Lecoanet et al., 2016; Cullen & Dehnen, 2010).

On the convergence of the rotation rate, in addition to the similar angular momenta of 10^7 and 10^8 throughout time, the rotation period encouragingly changes monotonically with higher resolution and by less with each increase. So we interpret the various quantities shown in Fig. 3.15 as demonstrating a range of behaviour from the apparently well-converged rotation rate and unbound atmosphere mass by 10^7 particles, through the possibly converged unbound ice, to the clearly un-converged orbiting rock.

3.7 Conclusions

In the second half of this chapter, we revisited the study of the giant impact onto the young Uranus that may explain its spin and other strange features (§3.4), to demonstrate the tools presented in Chapter 2 and to test the numerical convergence of such simulations.

We find that even large-scale results such as the rotation rate are not converged with standard-resolution simulations of 10^5 and 10^6 particles. The overall behaviour is similar in all cases, but small variations in the debris that falls back after the initial impact have a significant effect on the post-impact planet and its rotation rate, which appears to be well-converged with 10^7 and 10^8 particles, but not fewer. Similar but mildly less certain convergence is seen for the masses of atmosphere and ice that are ejected from the system, while the low mass of rock placed into orbit has not converged at all by 10^7 particles.

Increasing resolution is only one important challenge for developing more realistic simulations. We have here used a simple implementation of SPH with a focus on

simply increasing the number of particles. Future studies must continue to test high resolutions with, for example, more sophisticated equations of state and improved SPH formulations with better treatment of issues such as material and density discontinuities.

We conclude that standard-resolution simulations with $<10^7$ SPH particles can fail to produce reliable results even for large-scale properties of a planetary system. 10^7 and 10^8 particles appear to pass the threshold of resolving the major processes in a giant impact. However, different collisions and other specific simulation outputs will depend more or less strongly on the behaviour of smaller structures, with correspondingly different requirements for convergence. The highly non-linear nature of giant impacts and the combinations of short- and long-term, localised and distributed processes prevent simple predictions for how many particles will be sufficient for a given result to converge.

Chapter 4

Atmospheric Erosion

Now that we can run simulations with many more particles, like those in Chapter 3, new topics become possible to study. One compelling application for higher resolutions is the modelling of thinner, lower density atmospheres than the of-order-10% mass envelopes like Uranus'. In this chapter, I run the first full, 3D simulations to test atmospheric erosion by giant impacts in this regime.

I discuss in §4.1 how one can motivate the study of giant-impact erosion for the understanding of Earth's history as a habitable planet or the evolution of the many different atmospheres hosted by the now-thousands of detected exoplanets. However, the aim of this initial project is more the general exploration of this almost uncharted parameter space. For example: what does the impactor actually do to remove atmosphere in different scenarios? How easy it is to erode some atmosphere as opposed to all or none? And how do these answers change for head-on, grazing, slow, or fast impacts?

In §4.2 I describe the suite of simulations and present the results in §4.3 to begin answering these questions, with conclusions in §4.4.

4.1 Introduction

Most rocky planets form from tens of up to Mars-sized embryos that collide with each other after accreting from a proto-planetary disk (Chambers, 2001). At the same time, planets grow their atmospheres by accreting nebula gas from their surroundings and by outgassing volatiles from their interior (Massol et al., 2016).

For a young atmosphere to survive, it must withstand radiation pressure of its host star, frequent impacts of small- and medium-sized impactors, and typically at least one late giant impact that could remove an entire atmosphere in a single blow (Schlichting & Mukhopadhyay, 2018). This raises the question of how each of these

dangers affects an atmosphere. In this chapter, we focus on the direct, mechanical consequences of a giant impact.

Our own planet is the perfect example, since we can both observe an atmosphere that has survived to the present day and be confident that a giant impact took place late in its evolution – creating the Moon in the process. Several different Moon-formation scenarios have been proposed and revised, but no simulations have yet resolved a crust, ocean, or atmosphere for the proto-Earth (e.g. Lock et al., 2018; Ćuk & Stewart, 2012).

Focusing on the atmosphere, the Earth's volatile abundances are remarkably different from those of chondrites (Halliday, 2013), which act as a record of the early Solar System. Specifically, nitrogen and carbon are depleted compared with hydrogen, which could correspond to losing N_2 and CO_2 with an eroded atmosphere while retaining H_2O in an ocean (Sakuraba et al., 2019). However, the isotope ratios match those of primordial chondrites. Hydrodynamic escape preferentially removes lighter isotopes, while impacts remove bulk volumes of atmosphere. This suggests that impacts (not necessarily giant ones) are the primary loss mechanism, driving fractionation by removing more atmosphere than ocean while preserving isotope ratios (Schlichting & Mukhopadhyay, 2018).

Furthermore, the relative abundances of the noble gases helium and neon in different-aged mantle reservoirs suggest that the Earth had global magma oceans generated by giant impacts on at least two occasions, with corresponding losses of the atmosphere (Tucker & Mukhopadhyay, 2014). Fractionation of xenon also indicates a complicated history of atmospheric loss and the importance of ionic escape in addition to impact erosion and hydrodynamic escape (Zahnle et al., 2019).

Looking further afield, we have recently learnt not only that Earth- to Neptune-mass exoplanets are common, but that they host a remarkable diversity of atmosphere masses (Fressin et al., 2013; Petigura et al., 2013; Lopez & Fortney, 2014). The stochastic nature of giant impacts makes them a strong candidate for explaining some of the differences between planets that would otherwise be expected to have evolved similarly. Irradiation and photoevaporation from stellar winds can significantly erode an atmosphere (Lopez et al., 2012; Zahnle & Catling, 2017), but not enough to explain the diversity of planets around dim stars, where it should be much less effective.

Previous studies of giant-impact erosion have used analytical approaches and 1D simulations to estimate atmospheric loss from a range of impact energies (Genda & Abe, 2003; Inamdar & Schlichting, 2015). The one-dimensional nature of these studies also means that little work has been done on grazing collisions, in spite of

the fact that these are more likely to occur. Some studies have investigated oblique impacts for much smaller (of order 10 km) objects (Shuvalov, 2009), in which case the erosion is only ever in the local region and the planet's curvature is negligible. Their results showed a strong increase in local loss for more-oblique impacts, which, as we saw in Chapter 3, is the opposite of the trend for giant impacts. The typical approach for giant impacts is to estimate the ground velocities induced by the impact to study how much atmosphere is blown away. This misses out the complex details of a collision that can mix, deform, and remake both an atmosphere and core. Any precise study of the consequences of a giant impact therefore requires full 3D modelling of the planet and atmosphere at the same time.

Two studies have recently begun to scratch the surface of this field. One with two simulations of a head-on collision (Liu et al., 2015) of large super-Earths, and another with highly grazing impacts that do not interact the planets' cores (Hwang et al., 2018). This leaves a serious gap in our understanding of the formation and atmospheric evolution of planets in and outside the Solar System.

4.1.1 Analytical and 1D Models

We use two of the small number of previous erosion studies for comparison with our 3D simulations, both for the resulting loss of atmosphere and for the shock waves caused by the impact. Genda & Abe (2003) used 1D models to simulate the reaction of the atmosphere to a shock from vertical ground motion. Their results for the local fraction of lost atmosphere, $X_{\rm local}$, are fit well by a simple linear function of the ground speed, $v_{\rm gnd}$, in units of the escape velocity, $v_{\rm esc}$: $X_{\rm local} = -1/3{+}4/3\ (v_{\rm gnd}/v_{\rm esc})$ (their Eqn. 17), which they conclude is largely insensitive to the initial conditions of the atmosphere.

Inamdar & Schlichting (2015) also performed similar 1D, Lagrangian, vertical-shock simulations, but extended them to include thicker atmospheres up to 10% of the solid mass of the planet. They also created a model for predicting the ground speeds caused by a giant impact, by treating the collision as a point-mass explosion. They assumed momentum conservation for the shock speed, which leads to the ground speed as a function of distance, l, from the impact point: $v_{\rm gnd} = v_{\rm imp}(m_{\rm imp}/M)[(l/(2R))^2(4 - 3l/(2R))]^{-1}$ (their Eqn. 17, with a typo of missing the $^{-1}$), where $v_{\rm imp}$ is the speed of the impactor and R and M are the radius and mass of the planet. By combining their speed estimates with the 1D local erosion model, they presented predictions for the global atmospheric mass-loss fraction as a function of the impactor speed and velocity for different atmosphere masses (their Fig. 5).

4.2 Methods

4.2.1 Initial Conditions

As a recognisable starting point, we consider an impact similar to a typical Moon-forming scenario, with a target proto-Earth of mass 0.887 $M_\oplus$ and impactor of mass 0.133 $M_\oplus$. Both are differentiated into an iron core and rocky mantle, constituting 30% and 70% of the total mass, respectively. We use the Tillotson iron and granite equations of state (EoS) to model these materials, as described in §2.2.1.

For the atmospheres, we use Hubbard & MacFarlane (1980)'s hydrogen–helium EoS, detailed in §2.2.2. An ideal gas would probably be sufficient for the smaller atmospheres, but larger ones stray into the more-dense regime that this EoS is designed to include.

The atmosphere is adiabatic, while the Tillotson iron and granite layers are given a simple temperature–density relation of $T \propto \rho^2$, chosen somewhat arbitrarily to produce a core temperature of ∼5000 K, similar to the Earth today.

We test a range of atmosphere masses on the proto-Earth, namely 10^{-1}, $10^{-1.5}$, 10^{-2}, and $10^{-2.5}$ $M_\oplus$, as the lowest mass that we might hope to resolve adequately with 10^7 particles. We repeat some simulations with 10^6 and $10^{6.5}$ particles. To be precise, the particle numbers stated in this chapter refer to the number of particles per Earth mass (the bare target plus impactor mass is 1.02 $M_\oplus$). Thus, the numerical resolution stays the same for simulations with different-mass planets. For example, a '10^7' simulation that includes a 10% $M_\oplus$ atmosphere actually contains a total of $\sim 1.12 \times 10^7$ particles. For most of the suite (see §4.2.2) we focus on the 10^{-2} $M_\oplus$ atmosphere.

To produce the profile for each atmosphere mass, the surface temperature is kept fixed at 500 K for simplicity, while the surface pressure is varied until the desired atmospheric mass is obtained. In other words, the inner two layer profiles are integrated inwards from the surface, as described in §2.3.1, then the atmosphere layer profile is integrated outwards, until reaching a negligible minimum density of 10 kg m^{-3}. Separately, the total radius is also iterated to obtain the 30:70 mass ratio of iron to rock.

4.2.2 Impact Simulations

We specify each impact scenario by the impact parameter, b, and the speed, v_c, at first contact of the impactor with the target's surface, as described in §2.5.1 and

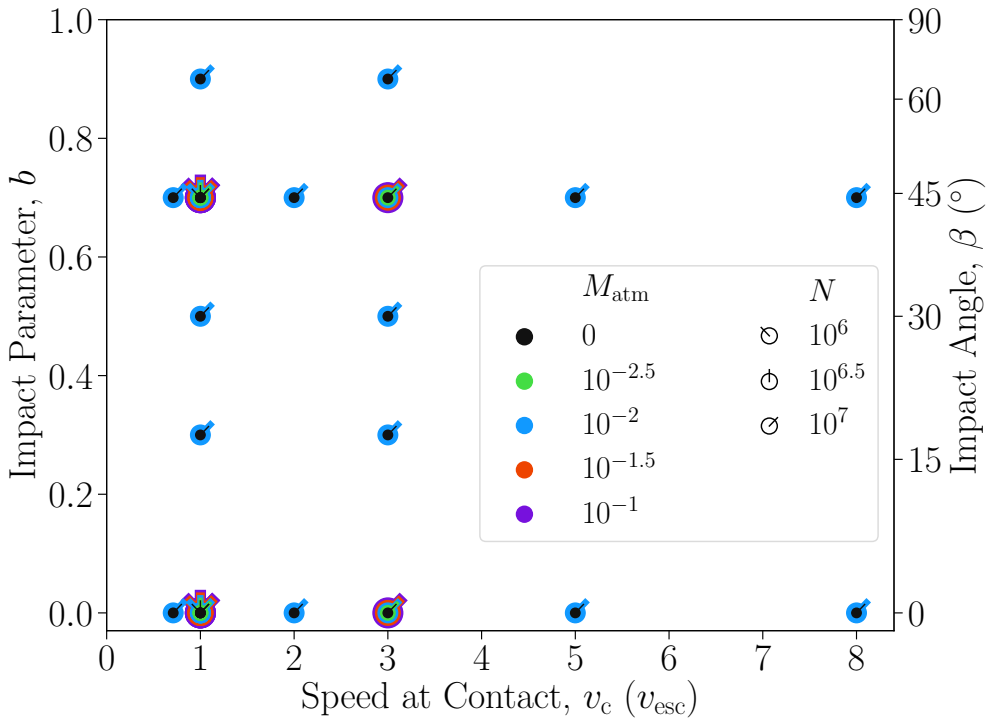


Figure 4.1: The suite of atmospheric erosion simulations, arranged by their speed and impact parameter at contact (see §2.5.1 and Fig. 2.5). As shown in the legend, the nested marker colours indicate the mass of the atmosphere (in Earth masses) for each simulation, while the line angles indicate the number of particles.

Fig. 2.5. The initial position of the impactor is set such that contact occurs 1 hour after the start of the simulation, to allow for some natural tidal distortion and not disrupt the system by suddenly introducing the large impactor right next to the no-longer-in-equilibrium target.

Fig. 4.1 shows the parameters for each simulation in the suite. Note that the $v_c = 0.75\ v_{esc}$ scenarios would require some third body to have slowed down the impactor during its approach to below the mutual escape speed. This is unlikely in the case of primary impactors falling in to the Earth in our solar system, but is a useful test for the consequences of a grazing impact resulting in a large bound fragment that will re-impact at a later time. It also lets us compare with other models, which predict little erosion in this regime.

At the high-speed end, given the Earth's position in the Solar System, 5 v_{esc} is around the highest typical velocity that might be expected for an impact (Raymond et al., 2009). For context, the Earth's orbital speed around the Sun is about 3 v_{esc}. The suite's extension to 8 v_{esc} both allows us to test the extreme end of the parameter space and is a regime that could be less unlikely in other planetary systems, for

example with a more massive star or a target planet deeper in the star's potential well. Furthermore, in studies like Inamdar & Schlichting (2015) where erosion is estimated as a function of the impactor's momentum, using very high velocities will allow us to test this degeneracy between impactor mass and speed across a wide range of momenta in planned future suites with different impactor masses. For the relatively small impactor mass used here, even 8 $v_{\rm esc}$ is not predicted by Inamdar & Schlichting (2015) to remove more than 3/4 of the atmosphere.

The simulations are run to an internal time of 100,000 s (roughly 28 hours) in a box of side 80 $R_\oplus$ to allow the tracking of ejecta. Any particles that leave the box are removed from the simulation. Throughout the first 10 hours we record snapshots every 100 s, for high time resolution during the impact and its immediate aftermath. To reduce data-storage requirements, we then output snapshots every 1000 s for the remainder.

4.3 Results and Discussion

We begin investigating the simulations with an overview of the general features and consequences of these classes of impacts. Then, we focus on the isolated effects of changing the impact parameter, speed, or atmosphere mass, and examine the time at which material is ejected. Finally, we consider the ground speeds and localised loss to compare the erosion results with previous estimates.

4.3.1 General Features of Impacts and Erosion

We choose four simulations to act as fiducial comparisons for the rest of the suite, demonstrating head-on and grazing, slow and fast scenarios. They stand out in Fig. 4.1 as the impacts for which we simulate multiple atmosphere masses. Snapshots from these fiducial simulations are shown in Fig. 4.2, for a target with a 1% $M_\oplus$ atmosphere, using $\sim 10^7$ SPH particles.

We can readily identify a variety of interesting differences between scenarios. At a basic level, in the head-on or slow cases, the impactor merges with the target, but for fast, grazing impacts, it may not. In addition to any differences in the resulting fraction of lost atmosphere, the timing and cause of loss can also vary significantly. For example, atmosphere may be eroded by – in approximate time order:

- Direct encounter with the very-much-not-a-point-mass impactor passing through, most dramatically demonstrated in the high-speed, grazing case ($4^{\rm th}$ row);

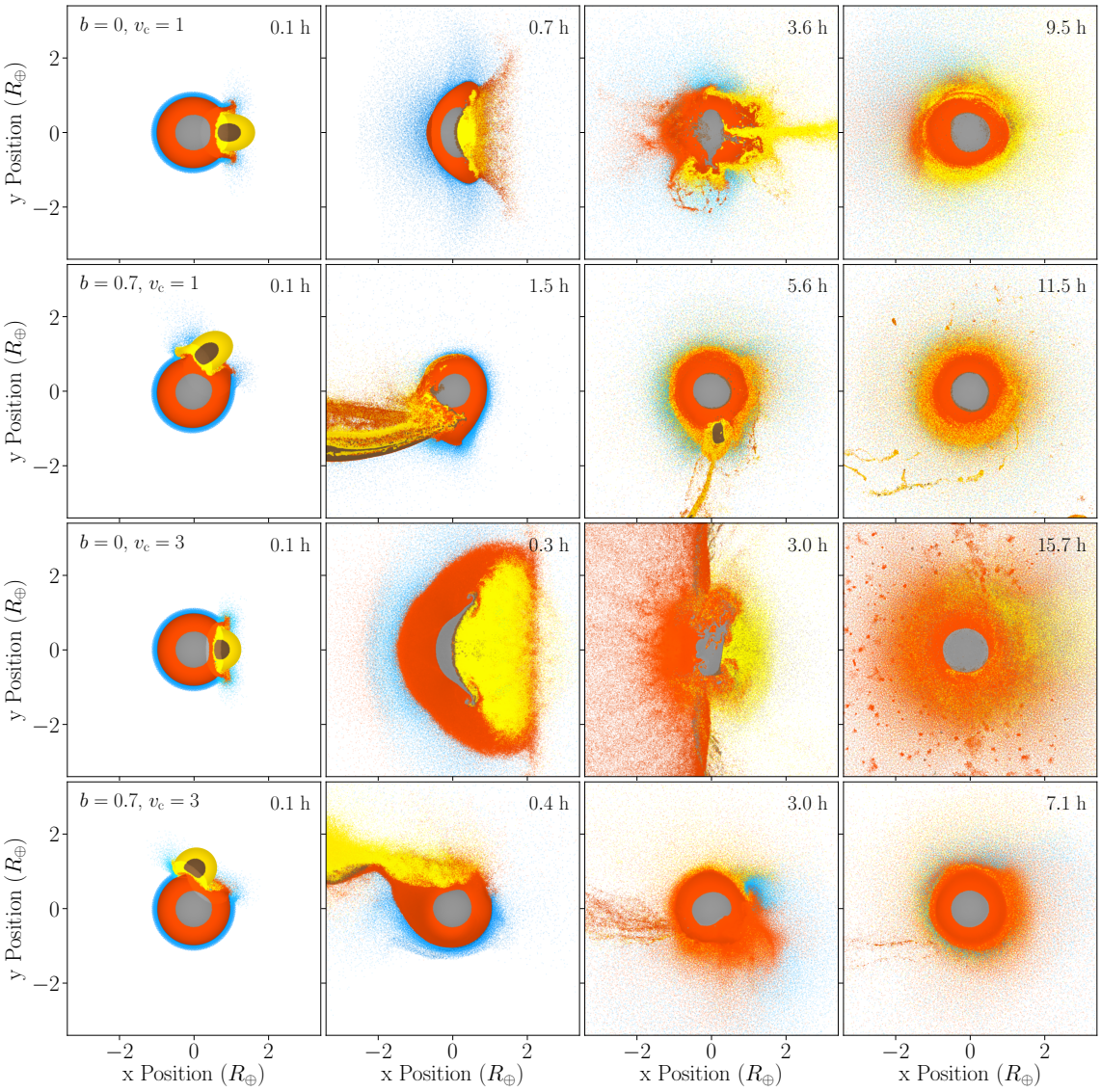


Figure 4.2: Illustrative snapshots from the four fiducial impact simulations with $b = 0$ and 0.7, and v_c (labelled throughout in units of v_{esc}) = 1 and 3, with the 1% $M_\oplus$ atmosphere and $\sim 10^7$ particles. Grey and orange show the target's core and mantle material respectively, and brown and yellow show the same for the impactor. Blue is the target's atmosphere. Note that the snapshots are at different times for each simulation to illustrate the evolution in each case. The impactors are travelling in the $-x$ direction at the moment they contact the target (see §2.5.1 and Fig. 2.5).

- The shock wave travelling through the planet from the impact point, which even erodes the mantle as well in the high-speed, head-on case (3rd row);

- Subsequent oscillations of the planet, such as the plume of impactor mantle in the 3rd snapshot of the low-speed, head-on case (1st row) – much like the large

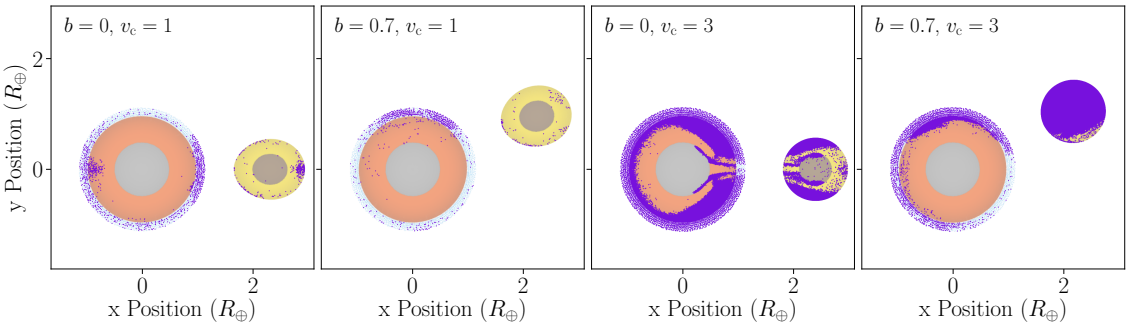


Figure 4.3: The particles that will become unbound and escape the system, highlighted in purple on a pre-impact snapshot, for the four fiducial impacts. The other particle colours are muted versions of those in Fig. 4.2 as a background for the highlighted ones. Only a thin cross-section of the particles that are within one smoothing length of $z = 0$ are shown for clarity.

splash created after dropping a stone into a pond;

- And the secondary impact of the impactor following an initial grazing collision, as in the 3$^{\text{rd}}$ snapshot of the low-speed, grazing case (2$^{\text{nd}}$ row).

All of these mechanisms may contribute to the total loss in a given scenario. This provides some context with which to consider the rest of the suite and some appreciation for the complexity created by all these processes intermingling.

As a brief aside, in the 3$^{\text{rd}}$ panel of the low-speed, head-on impact (1$^{\text{st}}$ row) in Fig. 4.2, in addition to the large plume we can see a blast of material on the $-y$ side of the target, in spite of the fact that this head-on collision should be rotationally symmetric. This material is the eruption of some impactor mantle that initially plunges deep into the target's core. Being much less dense than the iron core, it gets rapidly forced back out. The asymmetric direction in which it moves is a random consequence of the small deviations from perfect symmetry in our finite-particle representation of the system. Thus, for our simulations of the same impact scenario using different numbers of particles, the same eruption of material is produced at the same time, but with different orientations in the y–z plane. On the one hand, this highlights the imperfect symmetry of our SPH planets, which prevents the modelling of perfectly head-on collisions. On the other hand, this also demonstrates the importance of using fully 3D hydrodynamical simulations to study realistic giant impacts, where we should expect some level of asymmetry. At any rate, this feature ejects negligible unbound material, so does not affect the overall results of this specific study.

In mild contrast, the high-speed, grazing impact (4$^{\text{th}}$ row) in Fig. 4.2 is in some respects the most simple of the four. The impactor does remove some local atmo-

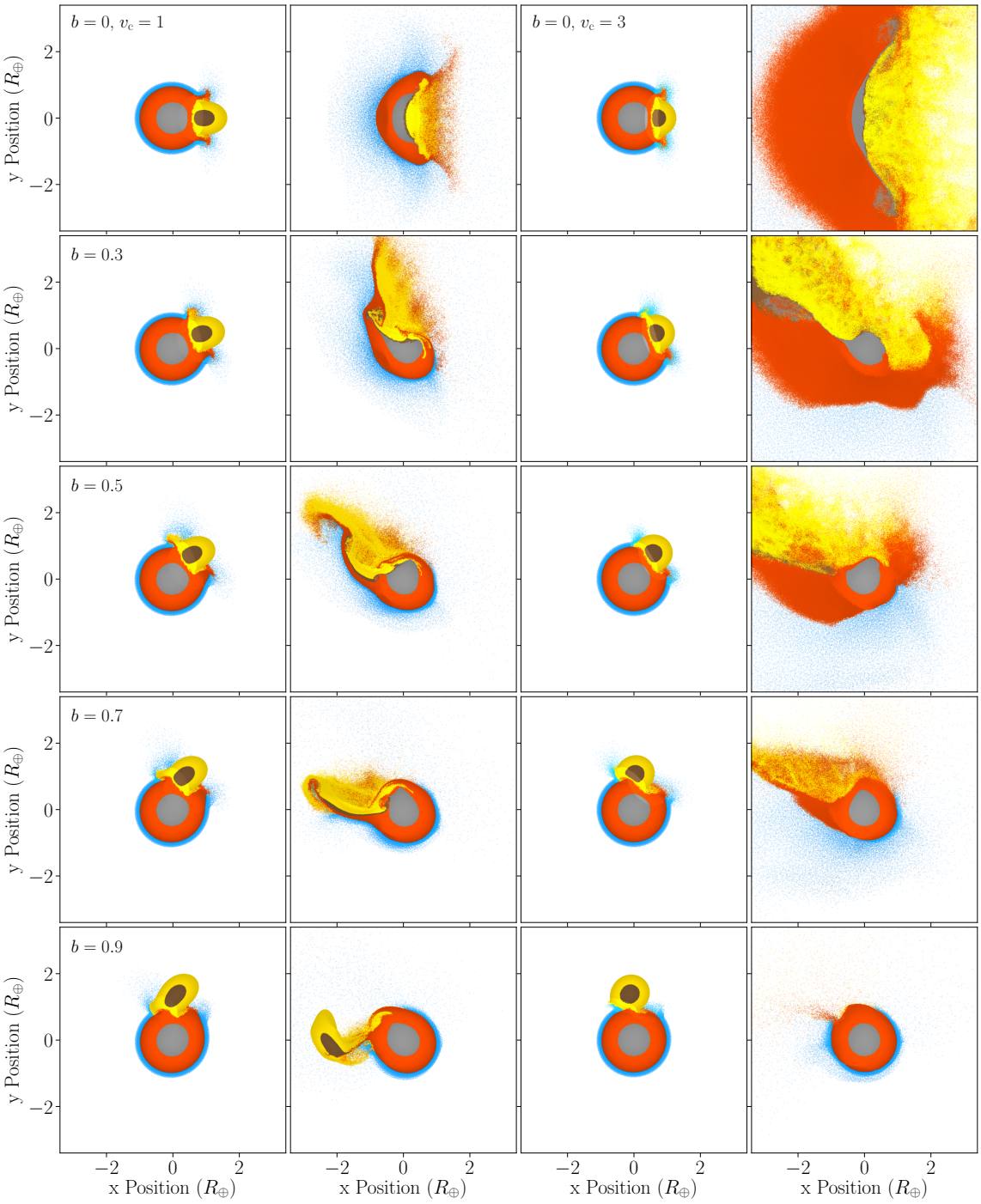


Figure 4.4: Pairs of snapshots ∼0.1 and 0.7 hours after contact for a range of impact parameters with a speed at contact $v_c = 1$ (left) or 3 v_{esc} (right), for the 1% $M_\oplus$ atmosphere and $\sim 10^7$ particles.

sphere that it encounters directly, but the majority of atmosphere globally is only ejected by the shock wave. Unlike the other three collisions, here the main body of the target is left relatively undisturbed, with little subsequent evolution while the mostly intact impactor rapidly exits the system.

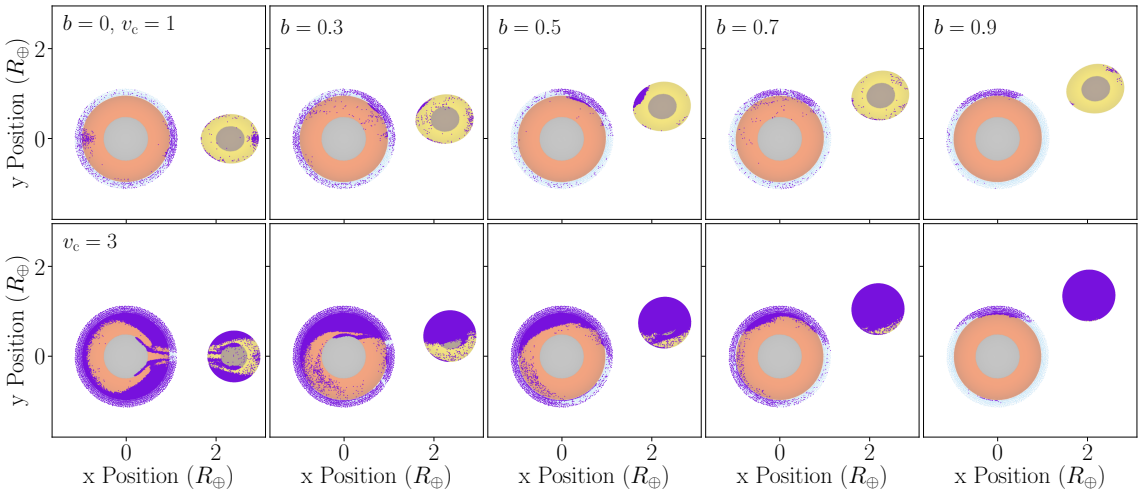


Figure 4.5: The particles that will become unbound and escape the system (see Fig. 4.3), for different impact parameters.

Fig. 4.3 highlights the particles that are eroded by these four impacts, selected by being gravitationally unbound and remaining so until the end of the 10^5 s simulation or until the time the particle exits the 80 $R_\oplus$-wide simulation box. The resulting mass fractions of lost atmosphere are 0.16, 0.13, 1.0, and 0.38, respectively. We revisit these final loss results in the context of the whole suite after presenting the rest of the simulations and introducing the previous analytical and 1D estimates for comparison.

For now, we see in Fig. 4.3 more or less the expected results from the above discussion of Fig. 4.2: the 1st case loses atmosphere around the impact point and the antipode; the 2nd shows less antipode erosion, suggesting a weaker shock, and primarily removed atmosphere in the direct path of the impactor; the 3rd has blasted off almost all the atmosphere and some mantle from the strong shock wave; and the 4th case is similar to the 2nd, but has taken some of the mantle in its path along with the atmosphere and blasted away some atmosphere around the antipode. The 4th impactor itself also remains unbound in this hit-and-run collision.

We now turn to the rest of the suite in a similar manner, continuing this initial overview of general behaviour. Fig. 4.4 shows pairs of snapshots at two early times from sets of changing-impact-parameter scenarios, with either the low or high fiducial speeds and the same atmosphere and number of particles. The particles that become lost in each case are highlighted in Fig. 4.5.

Filling in the gaps between the fiducial examples, we see a trend from global, presumably shock-driven erosion for low impact parameters, to direct, localised erosion for high impact parameters. At higher speeds, more of the grazing impactors can still deliver enough energy to drive some antipodal loss as well.

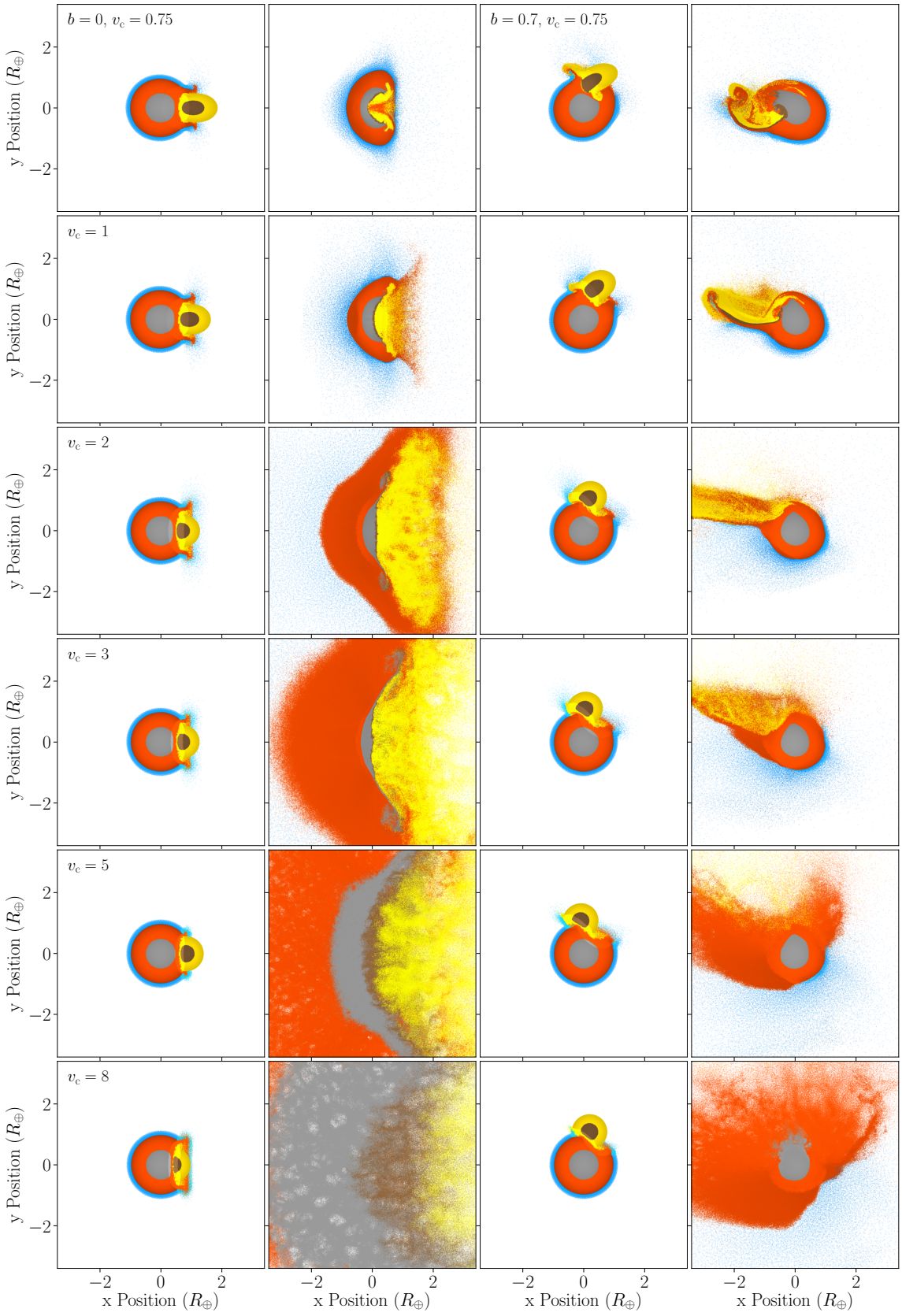


Figure 4.6: Pairs of snapshots ∼0.1 and 0.7 hours after contact for a range of speeds at contact (labelled in units of $v_{\rm esc}$) with an impact parameter $b = 0$ (left) or 0.7 (right), for the 1% $M_\oplus$ atmosphere and ∼10^7 particles.

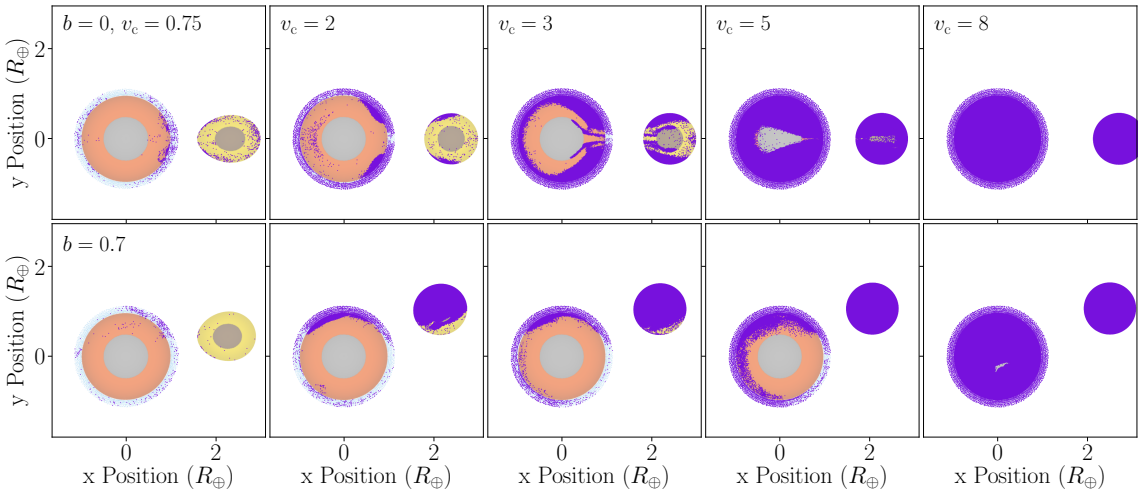


Figure 4.7: The particles that will become unbound and escape the system (see Fig. 4.3), for different speeds at contact (labelled in units of v_{esc}). The two $v_c = 1$ scenarios that also fit in these sets are shown in Fig. 4.5 and not duplicated here.

It is important to remember that these figures only show particles near the equator, defined as the $z = 0$ impact plane. For the head-on impacts, the system should be symmetric, but for grazing ones, higher latitudes may suffer less erosion. We return to this point in more detail in §4.3.3.

Changing focus to the speed at contact, Fig. 4.6 shows pairs of early snapshots from the two sets of changing-speed scenarios, with either the head-on or grazing fiducial impact parameters, with the lost particles highlighted in Fig. 4.7.

Even though the slowest impactors make contact at below the escape speed, they still erode some atmosphere. For head-on impacts, by $v_c = 2\ v_{esc}$, already almost all of the atmosphere is eroded. At higher speeds, more mantle is also lost, and $v_c = 8\ v_{esc}$ disintegrates the planet entirely. The grazing impacts remove systematically less than the head-on ones, and even by $v_c = 5\ v_{esc}$ with $b = 0.7$ some atmosphere still survives. $v_c = 8\ v_{esc}$ again destroys most of the planet.

Finally, we turn to different atmosphere masses, for the same fiducial impact parameters and speeds. Fig. 4.8 shows the lost particles in each case. Note that the speed at contact is always in units of the escape speed of the system, which is slightly faster for the more massive atmospheres. We see broadly similar results in all cases. For the slow, head-on impacts (1st row), the atmosphere significantly cushions the mantle from erosion for atmosphere masses above $\sim 10^{-2}\ M_\oplus$, at and below which the mantle erosion is similar to the case with zero atmosphere.

The low- and zero-mass atmosphere cases are also similar in the other three scenarios, although for slow, grazing collisions the atmosphere also affects the path of the

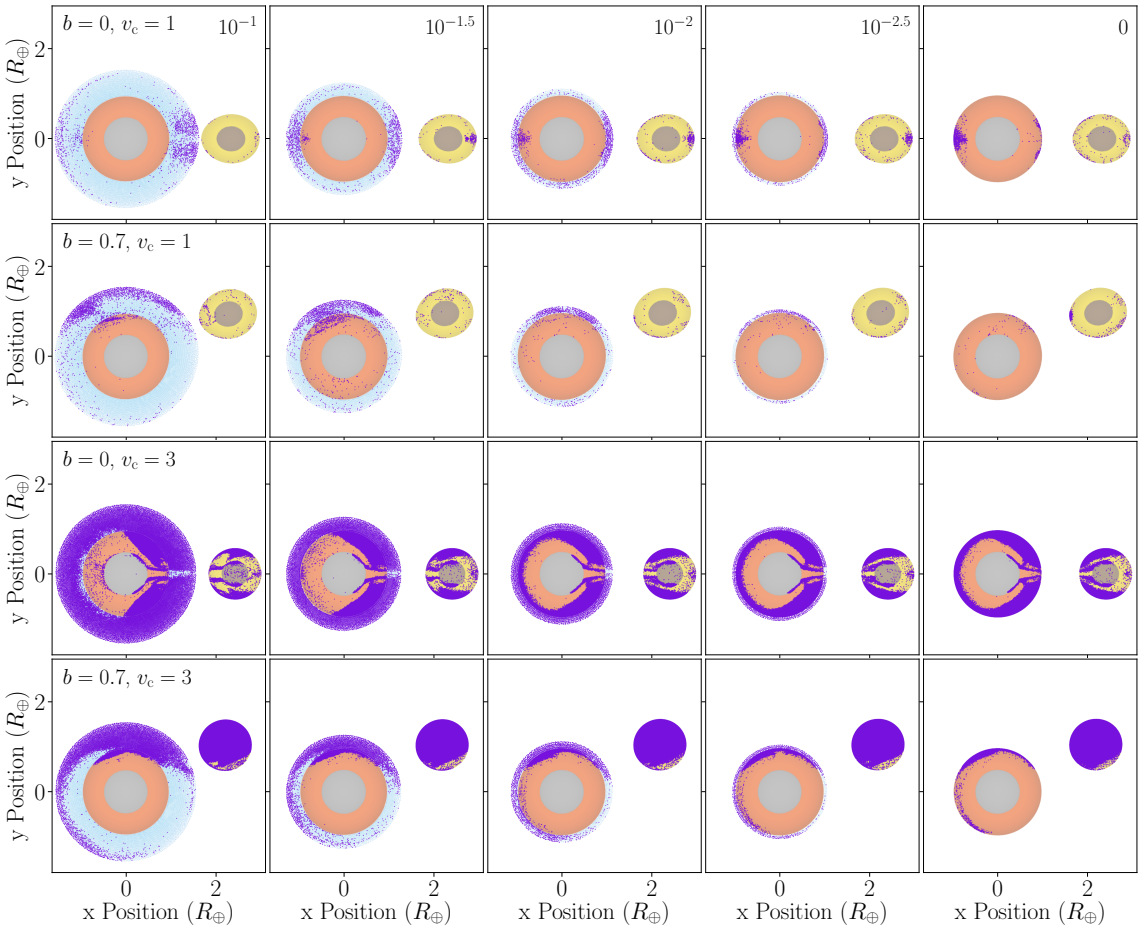


Figure 4.8: The particles that will become unbound and escape the system (see Fig. 4.3), for different atmosphere masses (top-right labels, in Earth masses) with 10^7 particles per Earth mass, for the fiducial impact scenarios of Fig. 4.2 (in which the atmosphere was 10^{-2} $M_\oplus$).

impactor as it passes through, making the comparison less direct. At higher speeds, the atmosphere mass makes less difference, especially in the head-on case. This is not surprising, since both any gravitational acceleration and hydrodynamical deceleration will have smaller effects.

4.3.2 Erosion Time Evolution

The time at which the lost atmosphere becomes unbound is shown in Fig. 4.9, for each of the changing-impact-parameter and changing-speed sets (as shown in Figs. 4.4 and 4.6). Significant atmosphere can be eroded after the initial impact, especially for slower collisions with low impact parameters. This corresponds to the potentially violent oscillations of the planet, shocking away surviving shells of atmosphere or even ejecting plumes of material, as seen in the 1st fiducial example (Fig. 4.2). For high impact parameters, delayed erosion can also be caused by the

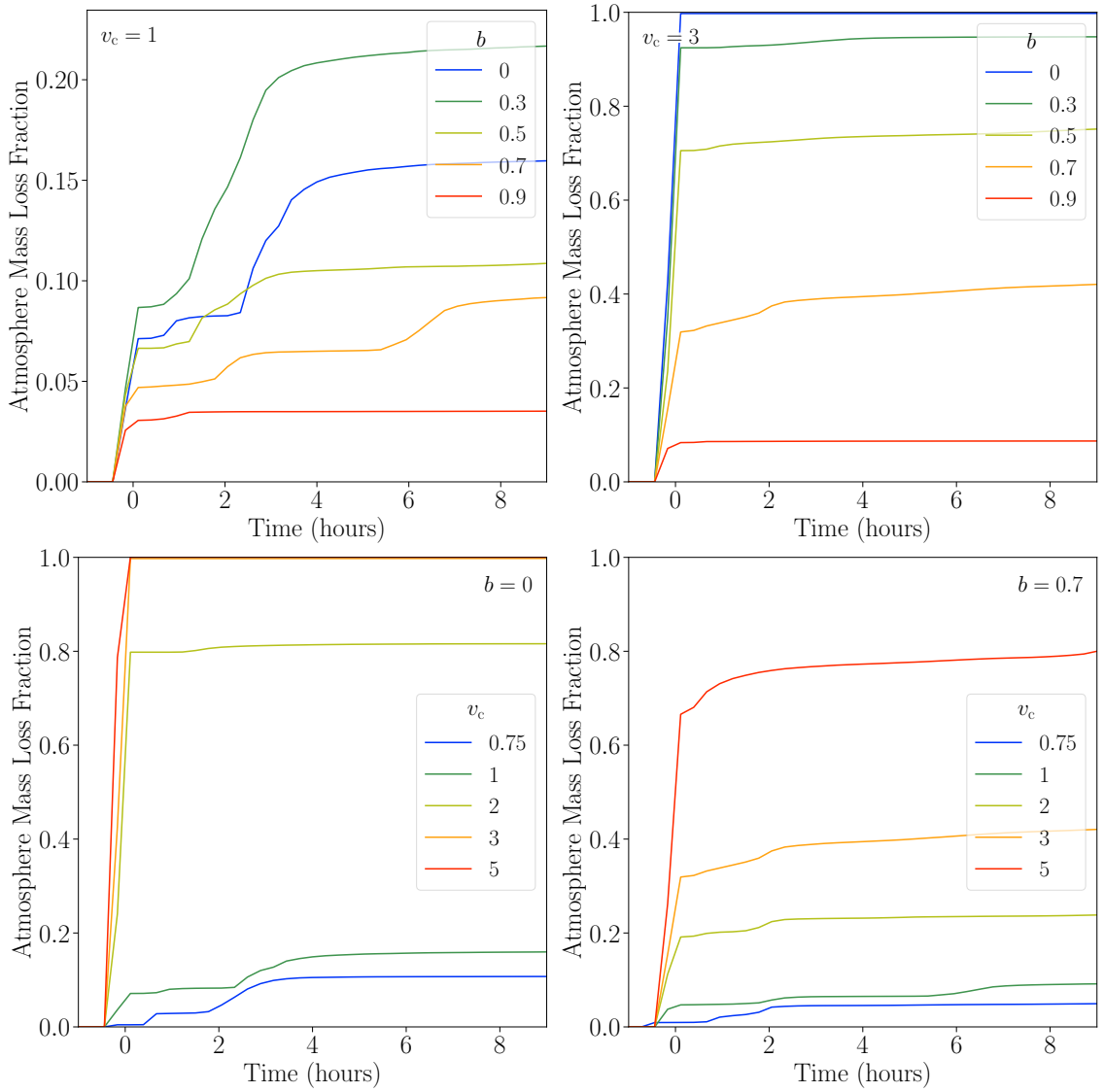


Figure 4.9: The early time evolution of the mass fraction of unbound atmosphere for different sets of impact parameters and speeds (labelled in units of v_{esc}) with $M_{atm} = 1\%\ M_\oplus$ and $\sim 10^7$ particles. i.e. the times at which the highlighted atmosphere particles in Figs. 4.5 and 4.7 become unbound. Note that the vertical axis in the top-left panel does not reach 1. Time = 0 is set to be the time of contact from §2.5.1, 1 hour after the start of the simulation.

secondary collision of grazing impactor fragments. However, given the low speeds required for a grazing fragment to return and the likely reduced mass of the fragment, this has a smaller effect.

In all cases, the majority of loss has finished by 4–8 hours after contact, and these lines remain flat to within a few percent up to the end of the 28 hour simulations. For impact speeds of $\sim 2\ v_{esc}$ or greater, the erosion is completed almost immediately,

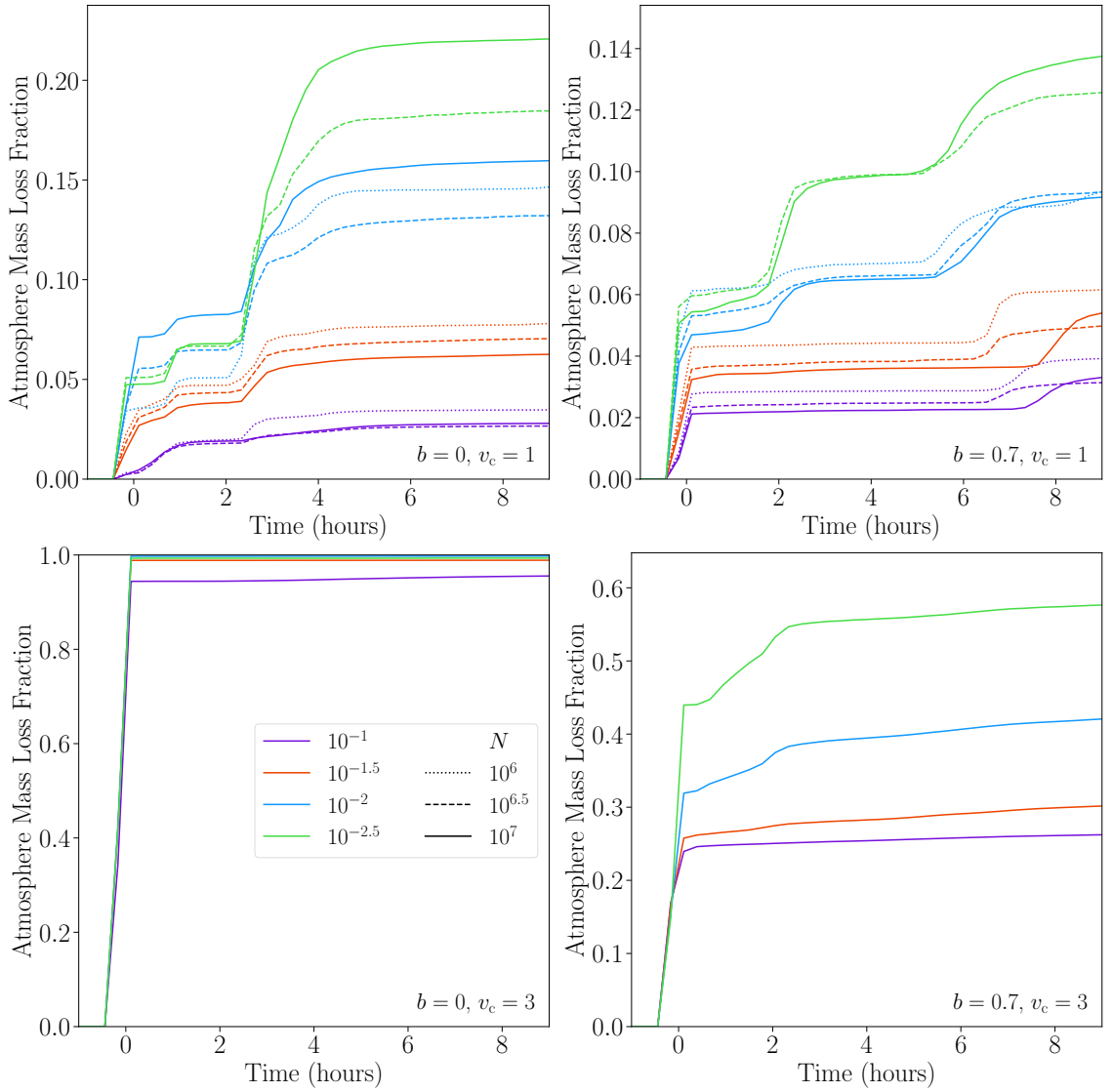


Figure 4.10: The early time evolution of the mass fraction of unbound atmosphere for the four fiducial impact scenarios with different atmosphere masses (labelled in the legend in $M_\oplus$ units). i.e. the times at which the highlighted atmosphere particles in Fig. 4.8 become unbound. Note that the vertical axes are not all equal. For the slower impacts, the dashed and dotted lines also show the loss evolution for lower numbers of particles, as given in the legend.

with little change after only the first 1 or 2 hours. For low impact parameters, this is simply because the entire atmosphere is blown away by the initial shock. For grazing collisions, it is the lack of re-impacting fragments that reduces any later erosion.

Fig. 4.10 shows the time evolution for the loss of different-mass atmospheres. The qualitative evolution is similar in most cases, especially for the 10^{-2} and $10^{-2.5}$ $M_\oplus$ atmospheres, and the total loss fraction is systematically lower for the thicker atmo-

spheres. As discussed with reference to Fig. 4.8, the drag of the atmosphere as the impactor passes through can reduce the erosion both immediately and by mitigating subsequent oscillations and secondary impacts. For the faster collisions, as before, the behaviour remains comparatively simple with more immediate erosion and the results are less affected by the atmosphere's mass in terms of timing.

This is also a good opportunity to test the results of using different particle numbers. We duplicated the two slower fiducial simulations with 10^6 and $10^{6.5}$ SPH particles (per Earth mass). For this initial project, we used 10^7 particles for the main suite to explore this new parameter space. From the second half of Chapter 3, we expect 10^7 to be approximately the minimum number required to resolve all the major processes in sufficient detail. However, for the atmospheric-erosion tests specifically, lower particles numbers still yielded results within 10% of the converged value.

We see similar results here with Fig. 4.9. The $10^{6.5}$ results agree to within ∼10% of the 10^7 head-on impacts for the larger two atmospheres and for all of the grazing impacts, and differ by <20% for the remaining lower mass, head-on cases. While this does not show complete convergence by $10^{6.5}$ particles, it is encouraging that even if the 10^7 particles were not entirely sufficient to resolve the erosion of 10^{-2} $M_\oplus$ atmospheres in the rest of the suite, we are unlikely to be wrong by more than ∼10%.

It is less clear if we should trust the thinner, $10^{-2.5}$ $M_\oplus$ atmosphere simulations with 10^7 particles, in which the atmosphere is only about 4 SPH particles thick. Based on the rough agreement of the $10^{6.5}$ particle results and the similarities with the 10^{-2} $M_\oplus$ atmospheres in Fig. 4.8, it seems likely that the localised regions of zero or total loss will be simulated correctly – as they probably could be with even a single layer of test particles that would simply react to the mantle movement below. However, regions where only some of the atmosphere should be lost are almost certainly not well enough resolved. Thus, we include the $10^{-2.5}$ $M_\oplus$ results in this chapter but treat the precise values as decidedly uncertain. Future simulations using 10^8 or more particles will be able to fully resolve these and even smaller atmospheres.

4.3.3 Ground Speed

The one-dimensional estimates of Genda & Abe (2003, hereafter GA03) predict the local atmospheric loss for a given vertical ground speed. By defining the 'ground' simulation particles as those in the outermost shell of the target's mantle, we can track their movement throughout the simulation as shock waves (and the impactor itself) perturb them, as illustrated in Fig. 4.11. The maximum radial speed at each location is given in Fig. 4.12 for the four fiducial simulations, and the times at which these peak speeds occur are shown in Fig. 4.13.

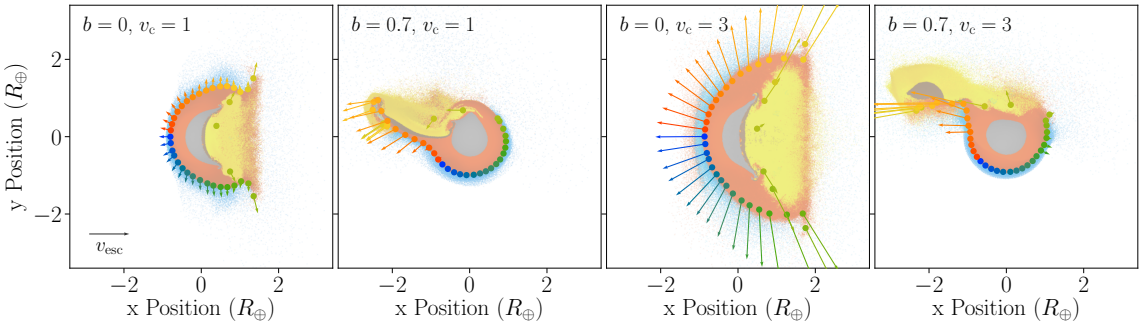


Figure 4.11: Example positions and velocities of the outermost 'ground' particles of the target's mantle, in the early stages of the four fiducial simulations. The particles are divided into sets of 10° longitude and, in this example, within ±5° latitude, to find the median positions and velocities. The colours show the particles' original longitudes, from 0° (pale green) at the point of contact, to 180° (red) and −180° (blue) at the antipode. The maximum speeds in Fig. 4.12 are taken from across all snapshot times, compared with the single snapshots shown here.

The two head-on impacts are symmetric in longitude (with zero set to be the point of contact) and show high peak speeds near the impact point and the antipode. For the slower of the two, the target recoils following the initial collision to shoot a plume of material back through the point of impact and a slightly less dramatic ejection at the antipode, causing the peak velocities in Fig. 4.12 at those longitudes. Some earlier erosion around the antipode is also caused by the initial shock wave, which is the origin of the maximum velocities at most of the other longitudes and latitudes. As shown in Fig. 4.13, this occurs about an hour before the peak recoil.

For the faster head-on collision, the impact is more destructive and no such bounce-back plume is seen. Instead, almost the entire surface is kicked immediately by the shock wave to faster than the escape speed, explaining the near-total erosion of atmosphere plus some lost mantle that was highlighted in Fig. 4.3. In both head-on cases, the speeds are lower at higher latitudes, which is simply a reflection of the rotational symmetry, evidenced by the equal peak speeds at equal times for ±90° longitude.

The two grazing collisions show similar behaviour to each other, with high speeds at positive longitudes, i.e. in the path of the impactor after it passes the point of contact. The rest of the planet is hit by the shock wave, but not one nearly as strong as in the head-on cases, and with only a mild peak at the antipode.

Unlike the head-on impacts, the grazing scenarios are not rotationally symmetric. Higher latitudes are less affected by the relatively small-size impactor and show little

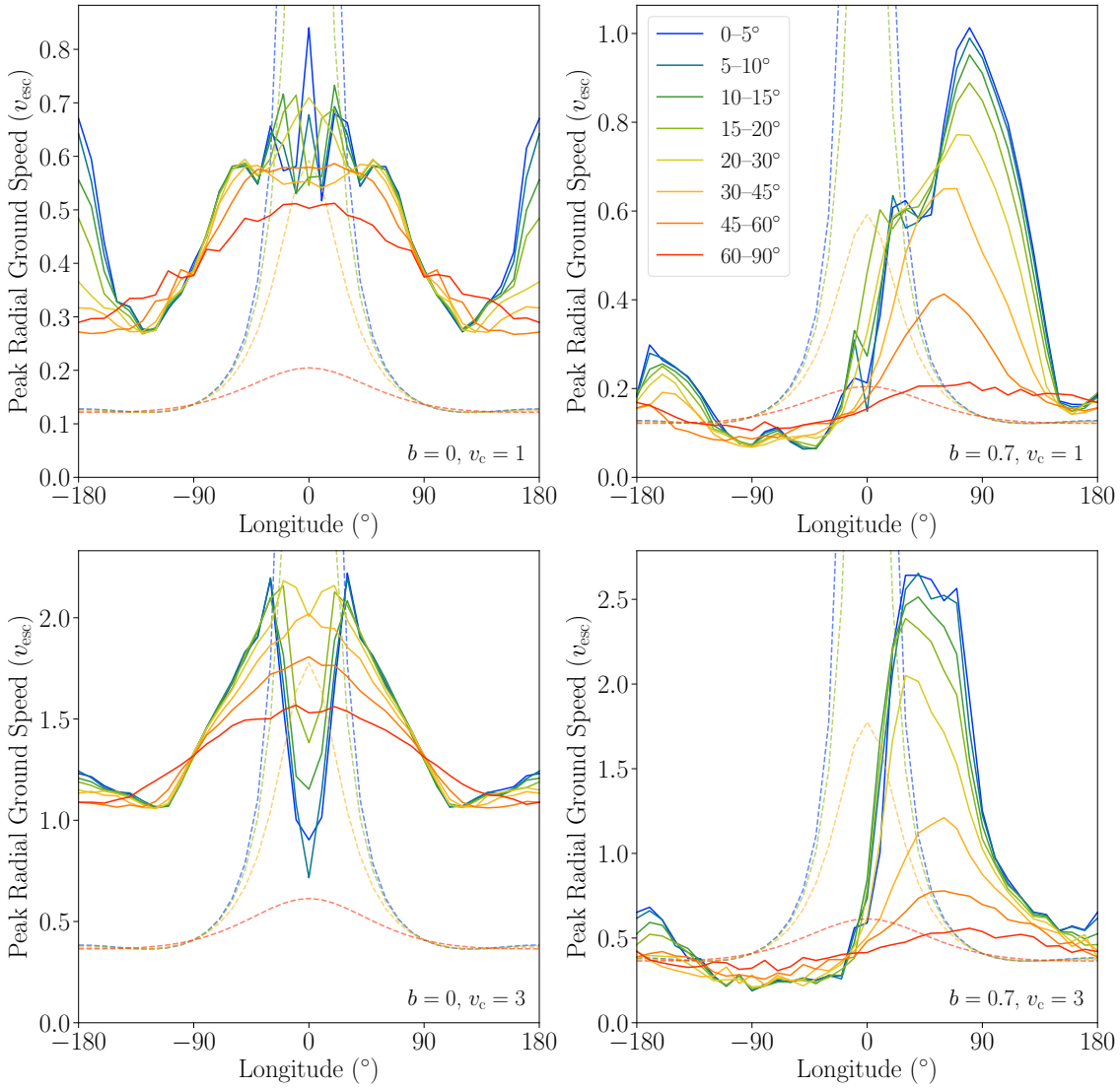


Figure 4.12: The maximum radial velocity of the outermost particles of the target's mantle as a function of longitude, in separate, approximately equal-area |latitude| bins, for the four fiducial simulations. The times at which these peak velocities occur are shown in Fig. 4.13. The dashed lines show the estimated ground speeds at some latitudes from Inamdar & Schlichting (2015), based on conserving a point-mass impactor's momentum in a spherical shock wave.

longitudinal variation.

In the slower grazing collision, the local loss around the impact site happens quickly, but the peak speeds everywhere else occur between one and two hours later, corresponding to the initial fallback of some impactor fragments and the recoiling oscillation of the planet. In the faster grazing case, the shock wave quickly produces the peak speeds across most of the surface, with little significant fall-back of

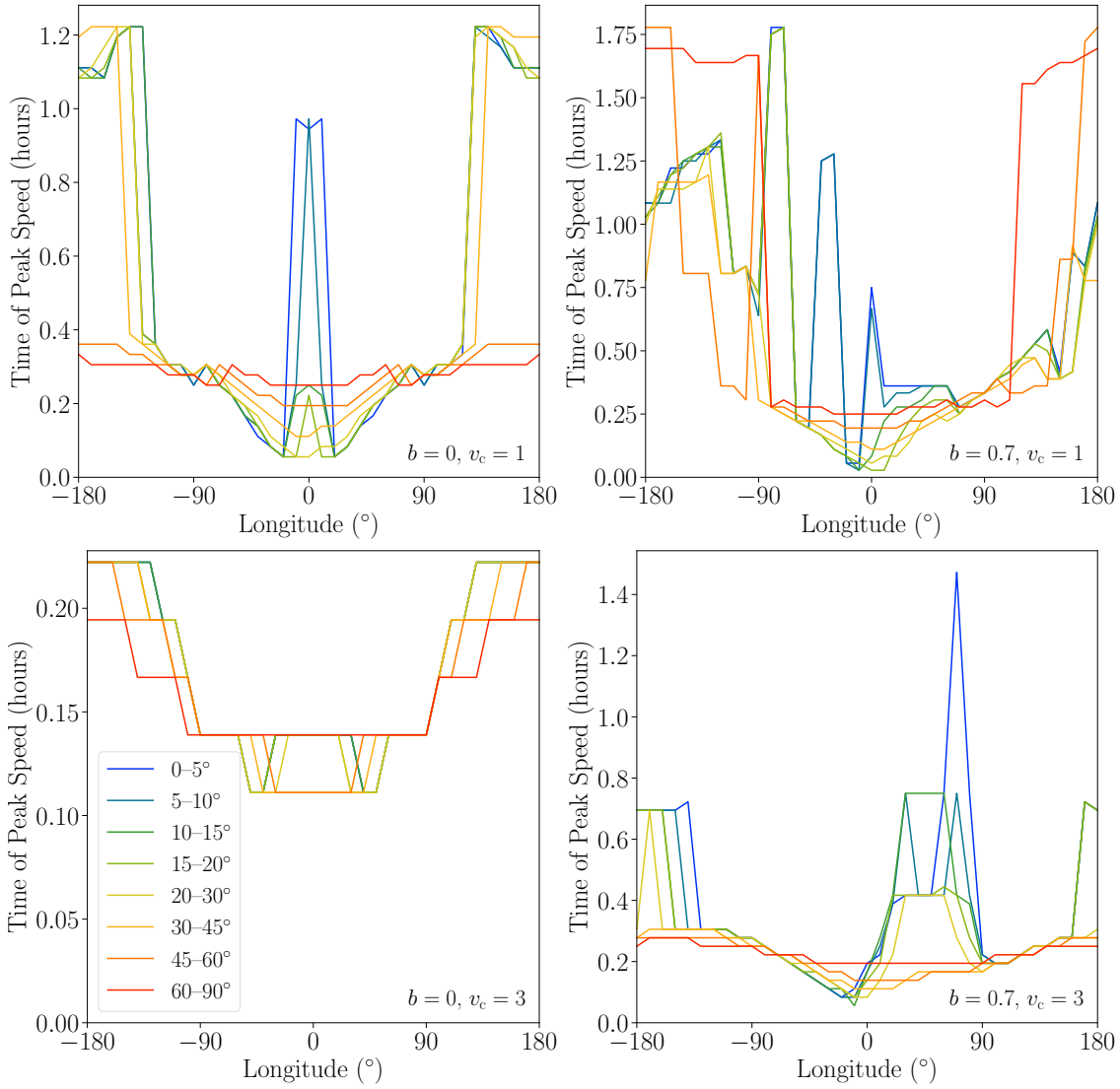


Figure 4.13: The times at which the peak ground velocities in Fig. 4.12 occur, as a function of longitude in separate |latitude| bins. The time 'resolution' between each snapshot is 100 s. The maximum velocities at low latitudes in the fast, head-on case occur at roughly the same times, so some lines are overlaid by the others.

fragments. The late times around the impact site are less meaningful since most of this material is carried away with the surviving impactor at a roughly constant speed slightly slower than the impactor's initial 3 $v_{\rm esc}$. The peak antipode speeds are caused by the violent sloshing of the target as it begins to resettle following the shock.

Also shown in Fig. 4.12 are the ground speeds predicted by Inamdar & Schlichting (2015, hereafter IS15). These are independent of the impact parameter so nominally correspond to head-on collisions. They assume that the impactor's momentum is transferred at the point of contact and is conserved as a spherical shock wave

propagates. While this therefore overestimates the ground speed close to the point-mass impact, it appears to significantly underestimate the peak speed everywhere else for head-on impacts. However, the speeds actually agree fairly well with the grazing simulations, especially away from the equator. This fits with our initial assessment of grazing (especially high-speed) impacts as being simpler, with a greater importance of shock waves through the relatively undisturbed target, compared with the more disruptive head-on scenarios.

4.3.4 Local and Global Atmospheric Loss

Now that we have examined the ground speeds across the planet for the fiducial impacts and introduced 1D and analytical estimates for comparison, we show in Fig. 4.14 the local mass loss in each region. The loss fractions broadly follow the distributions of peak speeds, with the trivial exception of the fast, head-on impact's location-independent total erosion, as expected from the global $>v_{\rm esc}$ ground speeds.

The GA03 results also match the simulated loss in many places, and quite closely for the faster impacts (precisely in the fast, head-on scenario, of course). Encouragingly, this implies that their 1D calculations and our SPH simulations reproduce the same results for a ground shock wave eroding the atmosphere above it once it arrives.

There are several reasons why this is not always the case for the more complicated scenarios we are dealing with here. Perhaps the most significant is that for these estimates we have taken a single value for the peak ground speed at each location, whereas in reality the atmosphere can be ejected at many points in time – as shown in Fig. 4.13. We also cannot fix this simplification by applying GA03's estimates at, for example, all local-in-time maximum ground speeds, simply because the atmosphere must still be present above the ground for a shock to remove it. After the initial impact, some parts of the atmosphere could survive relatively undisturbed and be removed by subsequent shocks. However, other parts could be partially shocked away to fall back down at a later time, which may or may not coincide with later shocks. Thus, the assumption of a single ground speed could either over or underestimate the actual local loss.

Another important issue is the large size of the impactor and its complicated interaction with the target, compared with a simple point-mass explosion that would better produce loss just from ground shocks. Significant amounts of material can thus be ejected directly by the impactor ploughing through the atmosphere and mantle, especially in grazing impacts. This ties in with the assumption that the ground speed is vertical, which is often not true for any regions around the path the impactor takes.

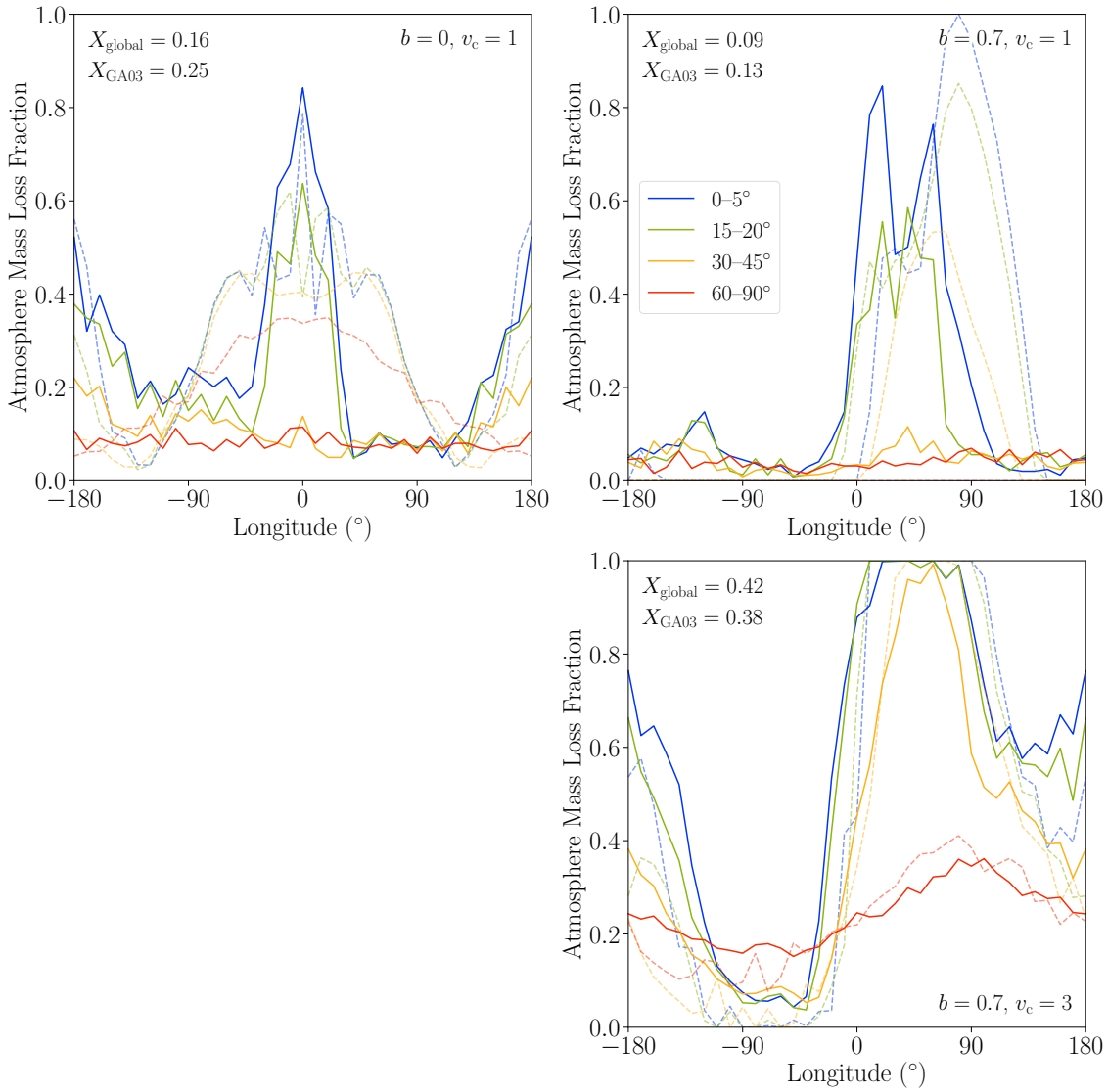


Figure 4.14: The loss fraction of local atmosphere (solid lines) as a function of longitude, in separate |latitude| bins, for the fiducial simulations – missing the head-on, high-speed scenario, which has uninteresting $X \approx 1$ everywhere. The dashed lines show the corresponding loss estimates from Genda & Abe (2003, GA03) using the peak ground speeds from our study that are shown in Fig. 4.12. Only a subset of latitude bins are shown for clarity. The top-left text shows the global loss fractions from the simulations and the GA03 estimates.

Finally, there are the underlying assumptions made and discussed by GA03, such as their use of an ideal gas EoS and ignoring lateral motion of the atmosphere, both of which are likely to be more valid in their targeted regime of even thinner atmospheres. However, the fact that our simulations agree with theirs in many cases suggests that these simplifications are often not too important. Other assumptions that both they and we make, such as neglecting radiative cooling in spite of the high

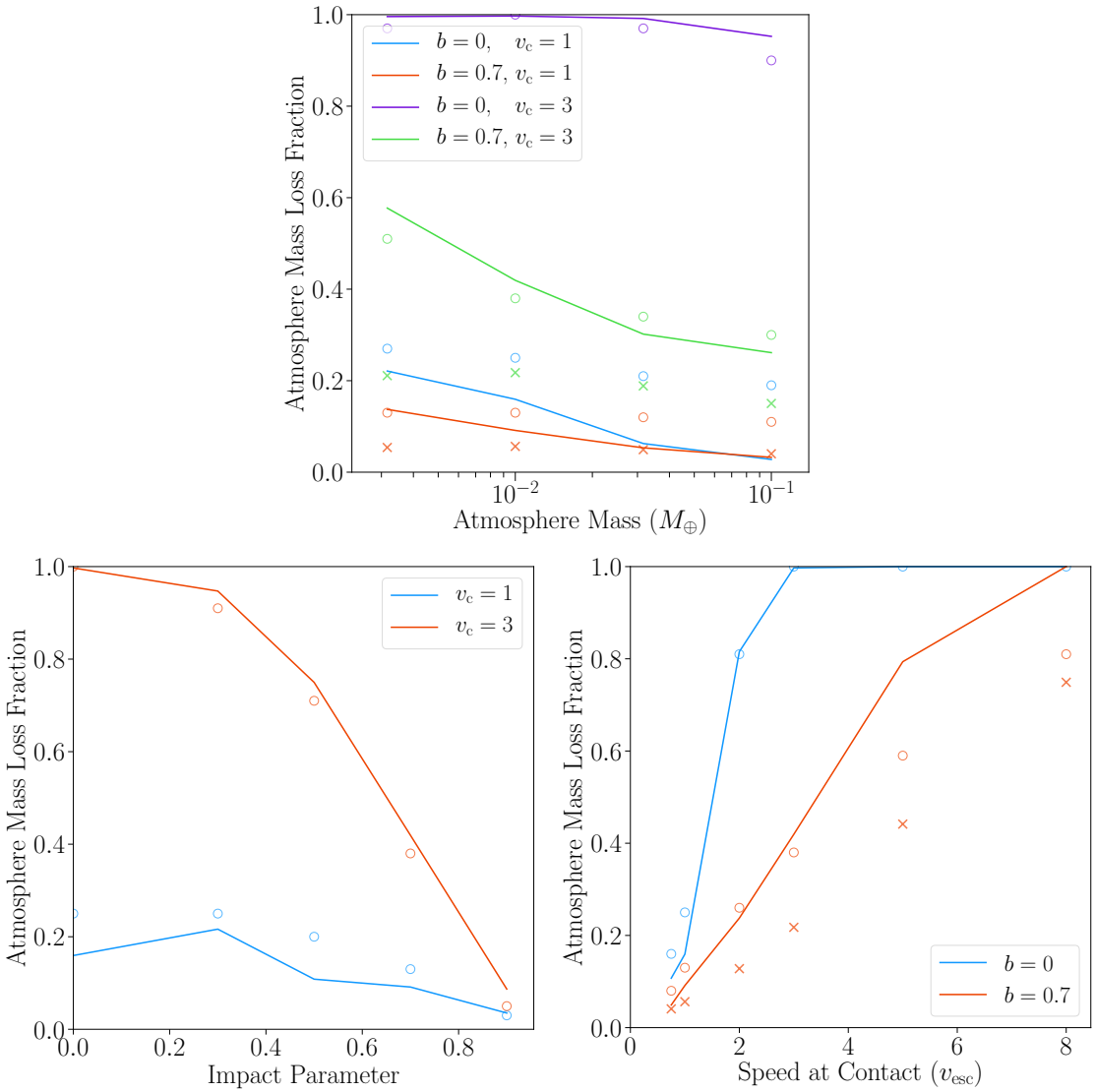


Figure 4.15: The lost mass fraction of the atmosphere for different: (top) atmosphere masses, in each of the fiducial impact scenarios; (bottom-left) impact parameters, for each fiducial speed; and (bottom-right) speeds at contact, for each fiducial impact parameter; all with $\sim 10^7$ particles. The circles show the corresponding Genda & Abe (2003) estimates based on the peak ground speeds. The crosses show the Inamdar & Schlichting (2015) estimates based solely on the impactor's mass and speed relative to the target.

shock temperatures, may still be relevant.

The overall results for the suite are presented in Fig. 4.15, showing how the fraction of lost atmosphere varies with atmosphere mass, impact parameter, and speed – the same subsets examined in previous figures. We can summarise the major trends with the uncontroversial conclusions that more atmosphere is usually lost from smaller

atmospheres, more-head-on collisions, and higher speeds.

However, for slower collisions, the loss is actually not a monotonic function of impact parameter, and a head-on collision does not cause the most erosion. By hitting slightly off-centre, the impactor can both deliver a strong shock through the planet while also encountering and eroding more atmosphere directly. Although more-grazing impacts can directly remove even more local material, they fail to deposit enough energy into the shock to erode as much atmosphere on the far side.

Apart from this, by following the same ground-speed analysis as for the fiducial impacts, the GA03 estimates continue to reproduce these results in most cases. As indicated by the ground speeds in Fig. 4.12, the IS15 estimates predict far less loss than most head-on collisions, but do better in comparison with grazing ones. The discrepancy with head-on collisions at low speeds is largely due to the fact that the peak velocities from the simulations are often not from the immediate shock wave. The behaviour around the impact site is also very different, as might be expected for a large impactor compared with a point explosion.

IS15 also assumed a constant-density planet for the shock wave to traverse, compared with the more complex and differentiated profiles here. More importantly, the internal structure of the planet changes dramatically as the large impactor plunges messily through the mantle, regardless of its initial state. At high speeds, the impactor can even reach the core of the target well before the shock wave has reached the other side.

For grazing collisions, the impactor interacts less with the target and the shock wave propagates to the far side more cleanly, so it is encouraging that our speeds and theirs tend to agree in this regime. This suggests that conservation of shock momentum is a valid assumption – provided that the shock is generated quickly and cleanly from a localised origin, without significant disruption of the rest of the planet. However, in terms of the total erosion, IS15 of course miss out on the direct effects of the impactor, which become more important for otherwise-simpler grazing impacts.

We might expect the loss fraction from a given impact scenario to converge with thinner atmospheres, meaning that we could extrapolate the 10^{-2} or $10^{-2.5}$ $M_\oplus$ results to smaller masses. This is supported by the good agreement of the GA03 estimates, which assumed a much thinner atmosphere than ours along the lines of the Earth's present-day, $\sim 10^{-6}$ $M_\oplus$ atmosphere. For the fast, grazing impact set in the top panel of Fig. 4.15, this is not seen for the $10^{-2.5}$ $M_\oplus$ atmosphere. However, as mentioned in §4.3.2, these smallest atmosphere simulations are pushing the limits of what 10^7 SPH particles can resolve, making the precise results for atmospheres below 10^{-2} $M_\oplus$ unreliable.

4.4 Conclusions

This chapter presented the first 3D simulations of giant impacts onto planets with thin atmospheres. We explored a wide variety of speeds and impact angles, as well as a range of atmosphere masses.

Several different processes can dominate the atmosphere loss in different scenarios, depending on, for example, whether the impactor can deliver a strong shock wave to remove atmosphere on the far side, or whether impactor fragments fall back after the initial collision. The interplay of these and other processes affects the total fraction of eroded atmosphere, the local distribution of where atmosphere is lost, and the time at which it is removed.

In general, we verified that a wide range of giant impacts can remove significant fractions of an atmosphere. For head-on collisions, there is a rapid change with increasing impact speed from very little erosion to total loss. However, for grazing impacts with changing speed – or for fixed speeds with changing impact parameter – there is a much more gradual transition that also displays complex, non-monotonic behaviour at low-to-medium impact angles.

By tracking the ground movement throughout the simulations, we compared these 3D results with Inamdar & Schlichting (2015, IS15)'s analytical estimates for the propagation of shocks from a giant impact, Genda & Abe (2003, GA03)'s 1D models for local shock-driven erosion, and IS15's combined predictions for the global loss in a given scenario. IS15's velocities significantly underestimate the maximum speeds in head-on impacts due to the significant deformation of the planet and violent post-impact oscillations, but agree fairly well with less messy, grazing impacts. For the same reasons, their global predictions underestimate the total loss but come within a factor of two for most grazing scenarios. Using our simulated ground speeds, GA03's estimates match the localised loss fractions well in most cases, especially when the direct encounter of the impactor with the atmosphere is not too important.

In the context of the Earth and the (canonical) Moon-forming impact, only around 10–20% of the atmosphere would have been lost from the immediate effects of the collision. This caveat of 'immediate' is important, because we have here only considered the direct, mechanical consequences of a giant impact. As examined by Biersteker & Schlichting (2019), the thermal effects of a giant impact heating the surface can erode more atmosphere than the ejection by shocks. In addition, we took the simple approach here of defining 'lost' atmosphere by particles that become gravitationally unbound. However, significant amounts can be ejected far from the planet while staying bound. In a real solar system, whether by interaction with the

solar wind or by leaving the target's Hill sphere of gravitational influence, much of the eroded-but-bound atmosphere could still be lost.

As a separate point, Genda & Abe (2005) showed that the presence of an ocean can significantly enhance atmospheric loss, such that in the canonical Moon-forming scenario, perhaps closer to half the atmosphere would be immediately removed.

In conclusion: the details of atmospheric erosion by giant impacts are complicated, and much remains to be studied. These first simulations form an exciting new base from which to explore the vast parameter space in greater detail. Promising targets for future projects include: filling in the gaps in the parameter space we have started to explore, in order to define different classes of giant impacts by their consequences and to study the transitions between them; and investigating how the erosion changes for different impactor and target masses, in part to determine what scaling laws may or may not exist with increasing the total system mass.

The results from studies like these could then inform models of planet formation to reveal the role giant impacts might play in creating the diversity of exoplanet atmospheres, and to more precisely predict the consequences of the Moon-forming impact for Earth's early atmosphere.

Chapter 5

The Lunar Argon Exosphere

In this chapter, we turn away from giant impacts themselves to study a particular feature hosted by the most famous outcome of a giant impact, the Moon; specifically the lunar argon exosphere with its links to the Moon's interior. This is just one example of a connection between the consequences of giant impacts and their role in planet formation to observations of present-day phenomena, motivated by unanswered questions raised by the LADEE mission four years ago.

The background for this project is described in §5.1. §5.2 contains the details of the data set used and an overview of the different aspects of the lunar argon exosphere that the available data can constrain. Our model is presented in §5.3. The results and their implications for the source, sinks, and regolith interactions of argon atoms are described in §5.4, and I summarise our conclusions in §5.5.

5.1 Introduction

The Moon possesses our nearest example of a surface-bounded exosphere, the most common type of atmosphere in the Solar System. As the atoms constituting an exosphere do not interact with one another during their ballistic trajectories over the surface, different species form independent systems. Their exospheric densities and variation with local time depend upon the sources, sinks, and surface interactions for that particular species. Hence, studying the lunar exosphere has the potential to teach us about the solar wind, the lunar interior and outgassing, the efficiency of volatile sequestration in polar cold traps, and the kinetics of adsorption and desorption in low-pressure environments (Stern, 1999; Watson et al., 1961; Wieler & Heber, 2003).

Argon is a particularly well-studied species in the lunar exosphere, having been first detected by the Lunar Atmosphere Composition Experiment (LACE), which

measured the $^{40}Ar/^{36}Ar$ ratio at the surface to be approximately 10 (Hoffman et al., 1973). This implied that the more important source of argon was radioactive decay of ^{40}K to ^{40}Ar, rather than solar-wind derived ^{36}Ar. The LACE results showed that the argon exospheric density decreased through the night and had a rapid increase that began just before sunrise, typical of a condensible gas that adsorbs to the cold nighttime surface and desorbs at dawn (Hodges & Johnson, 1968). In addition to this daily variation, there was a longer term decrease by a factor of ∼2 seen during the nine lunar days of observations (Hodges, 1975).

The Lunar Atmosphere and Dust Environment Explorer (LADEE) orbital mission produced a wealth of data concerning the lunar exosphere at altitudes from 3–140 km (Elphic et al., 2014). As well as measuring the daily and long-term variations in argon density during its 5 month mission, the Neutral Mass Spectrometer (NMS, Mahaffy et al., 2014) also determined the vertical structure of the exosphere and the variation with selenographic longitude. This led to the discovery that there was an enhancement in the argon exospheric density over the western maria, dubbed the argon 'bulge' by Benna et al. (2015). The long-term variation in the argon abundance was ∼28% during the LADEE mission, much smaller than had been seen 40 years earlier by LACE over similar time periods. However, Hodges & Mahaffy (2016) noted that 'the absence of sensitivity-related tests of the Apollo 17 instrument allows the possibility that the 1973 results were in part artefacts.'

Different models of aspects of the lunar ^{40}Ar system have been created to help interpret the available data, both in terms of the outgassing rate from the surface and the corresponding sinks that are necessary to yield the measured exospheric densities. Hodges (1975) used the LACE data and Monte Carlo methods to simulate an argon exosphere to constrain both the source rate and the surface interactions. Grava et al. (2015) also employed a Monte Carlo technique to follow an initial injection of argon atoms through their lifetimes in the exosphere, concluding that approximately 10% of the area of permanently shadowed regions (PSRs, Mazarico et al., 2011) is needed to cold trap atoms in order to provide a sufficiently high loss rate to match the LACE long-term decline in argon exospheric density. If a continuous background source had been included in their model, then larger cold traps would have been required to deplete the exospheric argon density rapidly enough.

Using their model, Grava et al. (2015) suggested that long-term variations in the exospheric density can be ascribed to sporadic moonquakes. Benna et al. (2015) noted the possibility of tidal stress as the source of the LADEE variation. In contrast, Hodges & Mahaffy (2016) proposed that seasonal fluctuations in the total cold trap area are responsible for the smooth, mission-long variations in argon density

measured by LADEE (Benna et al., 2015).

More than one proposed explanation also exists for the bulge – the persistent enhancement of exospheric argon localised over the western maria. Benna et al. (2015) noted the similarity between the longitudinal variation in argon and the map of near-surface potassium returned by the Lunar Prospector Gamma Ray Spectrometer (LPGRS, Lawrence et al., 1998), suggesting a localised source. However, Hodges & Mahaffy (2016) asserted that the lifetimes of argon atoms in the lunar exosphere are too long for them to reflect their source locations. Instead, they suggested that the bulge results from lower desorption energies at these longitudes, which would cause argon to spend less time residing on the surface where it cannot be measured.

In this chapter, we develop a new model of the lunar argon exosphere using Monte Carlo methods. This approach is similar to those described by Smith et al. (1978), Hodges (1980b), and Butler (1997), in their studies of the helium and water exospheres of the Moon and Mercury. We apply our algorithm to address the questions of which – if any – of the proposed models could be responsible for the longitudinal and long-term variations in the argon densities measured by LADEE. Specifically, we produce the first simulations with: spatially varying surface interactions; a source that reflects the lunar near-surface potassium distribution; and seasonally varying cold trap areas.

5.2 Data

The NMS on LADEE measured the density of argon (and other species) in the lunar exosphere from 22nd November 2013 to 17th April 2014 at a wide range of altitudes, longitudes, and local times of day at latitudes within 30° of the equator. Derived data, including background-subtracted argon number densities at altitude, were obtained from the Planetary Atmospheres Node of NASA's Planetary Data System.

We apply two cuts to the entire LADEE argon data set to produce the subset of data used here. In the full data set, any densities that are negative after the background subtraction (due to noise) have had their values set to zero, which causes the mean to be artificially high if these are either included as zeros or discarded. We only use data in bins of local time of day and selenographical longitude for which more than half of the observations are positive. In this way, the medians of the resulting sets of measurements should not be biased by this prior treatment of negative values. Also, an unaccounted-for temperature dependence of the instrument background can affect the densities just after midnight (M. Benna, personal communication, 2016).

This problem is only important at this local time of day and for very low densities, where the instrument background was large compared with the signal. We therefore discard data at local times of day 180°–265° (where 0° is noon, 90° is sunset, 180° is midnight, and 270° is sunrise). Fortunately, the LACE data largely fill the overnight gap, and we supplement the LADEE measurements using the results presented by Hodges (1975).

Four complementary aspects of the argon exosphere can readily be studied: (1) the change in density with altitude; (2) the long-term variation in the global density during the months of LADEE's operation; (3) the density distribution with local time of day; and (4) the dependence on selenographical longitude, showing the bulge over the western maria (Benna et al., 2015).

5.2.1 Densities at Altitude

In order to study the final three of these distributions, we need to account for the measurement altitude varying from 3 to 140 km. Following convention, and for comparisons with the LACE data, we convert all measurements to the corresponding densities that would have been measured at the surface. To do this, we consider the expression derived by Chamberlain (1963) linking the number density in an exosphere as a function of height, $n(h)$, to the number density and temperature at the surface, n_0 and T respectively. For a spherical body with mass M and radius r,

$$n(h) = n_0 \exp\left[-\frac{GMm}{kT}\left(\frac{1}{r} - \frac{1}{r+h}\right)\right] , \tag{5.2.1}$$

where m is the mass of the particle, G is the gravitational constant, and k is the Boltzmann constant. This is the generalised form of the (isothermal) barometric law. We refer to this as a 'Chamberlain distribution'. The altitude dependence used by Benna et al. (2015), which results from assuming a constant acceleration due to gravity, is the first-order expansion of Eqn. 5.2.1 for small h.

The surface temperature varies as a function of latitude and local time, with a particularly rapid variation around the terminators. Lateral transport of molecules implies that the density at altitude will not reflect only the single sub-detector surface temperature, because particles reaching the detector will have originated at a variety of locations and temperatures. Therefore, we expect the real distribution to be a sum of many Chamberlain distributions for different temperatures, weighted by the number of particles that come from each one. As a practical model, we approximate this distribution with a sum of just two Chamberlain distributions at different temperatures and find the best-fit parameters at all times of day using our simulations, as detailed below in §5.2.1.1.

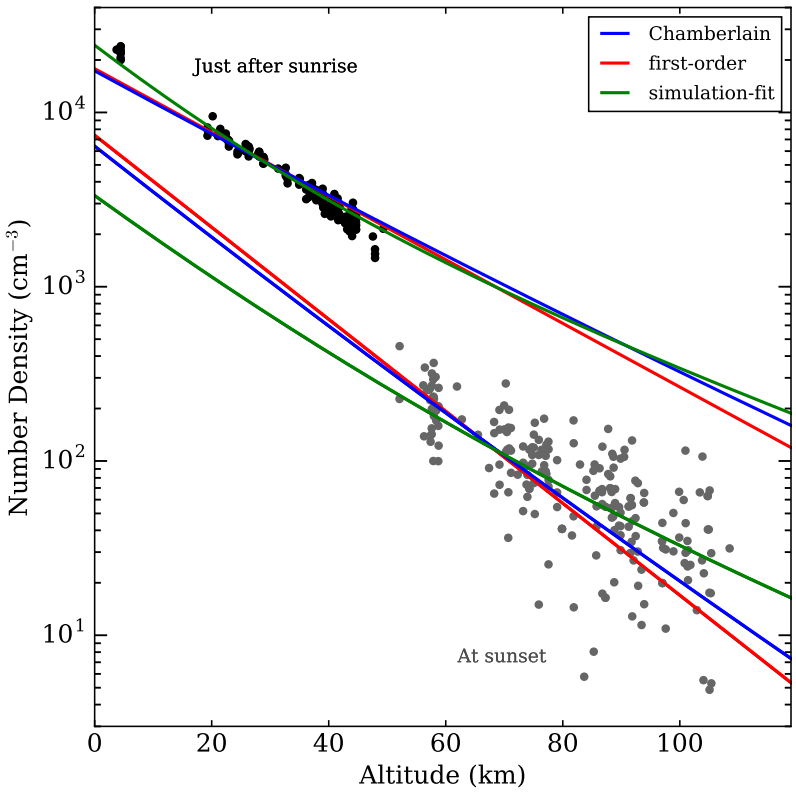


Figure 5.1: Two examples of the change of density with altitude near the terminators. The black and grey points show the LADEE data just after sunrise (273°) over the maria and at sunset (90°) over the highlands respectively. The coloured lines show the three models at the same local times of day, scaled in magnitude to match the data. The data have had the long-term variation extracted, as described in §5.2.2.

Fig. 5.1 shows two illustrative examples of the altitude variation of the LADEE data near the terminators, where the three models described above differ most from each other. It is apparent that the choice of extrapolation to zero altitude can significantly affect the inferred number density at the surface. Hurley et al. (2016) found a similar discrepancy between a Chamberlain profile and their model of the helium exosphere.

The LADEE data cover only a small range of altitudes at most local times of day. So, we use our simulations to test how accurately the three different models predict the simulated density at the surface at every time of day, from observations taken at the average LADEE altitude of 60 km. We found that our 'simulation-fit' model of two Chamberlain distributions with different temperatures successfully predicts the density at the surface to within 12% everywhere. This compares with overestimates as high as 337% and 416% for the single Chamberlain distribution and its first-order expansion respectively, which both use single temperatures (from the Hurley et al. (2015) model described in §5.3.4). These deviations are most pronounced near

the terminators. For example, just before sunrise, where the sub-detector surface temperature is very low, many particles will also be detected that originated at the hot surface after sunrise, which the Chamberlain and first-order models cannot account for. Away from the terminators, in regions where the surface temperature varies only slowly with local time of day, all three extrapolations correctly predict the simulation's density at the surface to within a few per cent, although the first-order model does less well at higher altitudes. We use the simulation-fit model to infer all LADEE argon densities at the surface reported in this chapter.

The agreement of the simulation results with the LADEE altitude data at all local times of day is also good evidence that our underlying models for the simple thermal desorption of argon atoms from the surface are appropriate. The data also do not show any clear differences between the altitude distributions above the mare and highland regions.

5.2.1.1 Altitude Model Fitting

The 'simulation-fit' model is a sum of two Chamberlain distributions with three free parameters: the relative amplitude of the two distributions, and their different temperatures. The total amplitude is also free but is simply set by the observed density at altitude.

In order to fit these parameters at all local times of day and latitudes, we first obtained the simulated density at equilibrium for a range of altitudes (from 0 to 140 km in 10 km steps). The best-fit parameters for each square degree bin were then determined by finding the minimum χ^2 for the simulation data, and are publicly available at icc.dur.ac.uk/index.php?content=Research/Topics/O13.

In our steady-state model, the argon density near the end of the lunar night is extremely low, especially at high latitudes and altitudes. The LADEE data in these same regions were discarded, as discussed in §5.2, so this was irrelevant for our analysis of the dataset. Thus, the best-fit parameters in these lowest-density regions were not examined in detail and may suffer from the fewer observed particles and high scatter in the simulation results.

This model is dependent on the argon distribution with time of day in the simulation. We repeated the analysis with a non-physical isotropic distribution, to quantify the maximum possible error that from discrepancies between our model and the real distribution. This amounted to ∼7% near the end of the night and below 1% elsewhere (within ±30° latitudes).

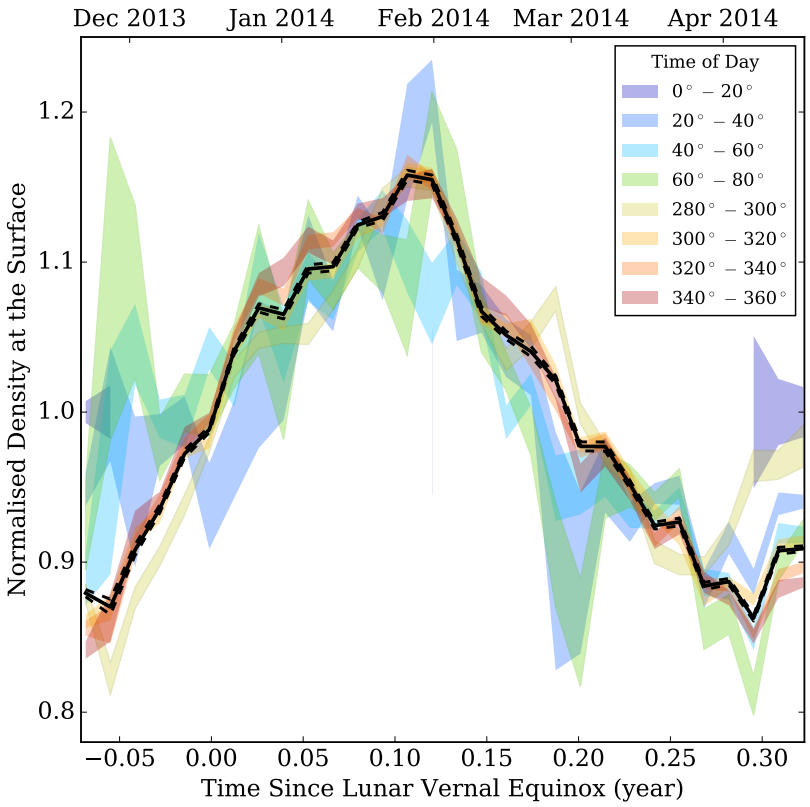


Figure 5.2: The long-term variation in the LADEE argon density, normalised by the mean density. The different colours show the median variation and its 1σ uncertainty at the corresponding local time of day, as given in the legend. The solid black line shows the total mean variation, weighted by the errors on the time-of-day values. Dashed black lines represent the 1σ uncertainty on this mean. Time is measured from the lunar vernal equinox on 17 December (Archinal et al., 2011).

5.2.2 Long-Term Variation

As the argon abundance varies dramatically with the local time of day, to determine the long-term variation we first split the LADEE data into bins of 20° in local time and calculate the mean density for each bin. The long-term variation is then calculated by subtracting the corresponding mean value from every measurement. To combine the data across all local times, we then divide by the same mean value to give the normalised deviation at a given long-term time. The median deviations are shown in Fig. 5.2, along with the weighted mean of all of these curves. The general agreement across the different times of day is notable, as is the relatively smooth variation.

The peak-to-peak change in density is 28%. Benna et al. (2015) noted the somewhat

similar magnitude and timescale of the variation to that observed by LACE, and offered that transient changes in the release rate of argon from the Moon's interior are a plausible cause of this variability. Hodges & Mahaffy (2016) instead suggested that it is part of a periodic fluctuation with a period of half a year. They proposed that this is driven by seasonal variations in the polar cold trap areas.

All subsequent figures in this section show densities at the surface with the mean long-term variation removed from the data. This is done by dividing each data point by the normalised long-term variation at the time of the measurement. This reveals what LADEE would have observed had the exosphere been in a steady state.

5.2.3 Local Time of Day

Fig. 5.3 shows the distribution of argon with local time of day from LADEE and LACE data, with the LADEE data corrected for altitude and long-term variability as described above. The figure shows just three representative examples across the mare and highland regions for clarity. The LADEE (and LACE) data show very similar behaviour at all selenographic longitudes. In particular, the timing of the sunrise peak is insensitive to the longitude, as highlighted by the inset panel, occurring at local times between $269° \pm 1°$ and $272° \pm 1°$. However, the argon density is greater in the maria than that in the highlands by a factor that ranges from a few tens of per cent up to almost a factor of 2 at sunrise. The LACE densities have been rescaled in amplitude to match the inferred surface density at sunrise from LADEE at the location of LACE, to extract the long-term variation between the data sets. Note that the lowest late-night LACE measurements may be below the instrument's sensitivity (Hoffman et al., 1973).

The large peaks in density around sunset and sunrise are both fed by particles migrating away from noon, where the higher temperatures mean larger hops. At sunset, the temperature is lower, so particles do not hop very far, but it is not yet cold enough to trap argon on the surface for long periods of time. Any particles that do stick to the surface at night will rotate with the Moon toward sunrise, creating the enormous peak when they warm up at dawn and reenter the exosphere. This peak extends back into the night because the particles fly in all directions, with a typical hop distance of a few degrees at post-sunrise temperatures.

5.2.4 Selenographic Longitude – The Bulge

The change of density with selenographic longitude is shown in Fig. 5.4 for different slices in local time of day. As was evident from Fig. 5.3, the density at all local

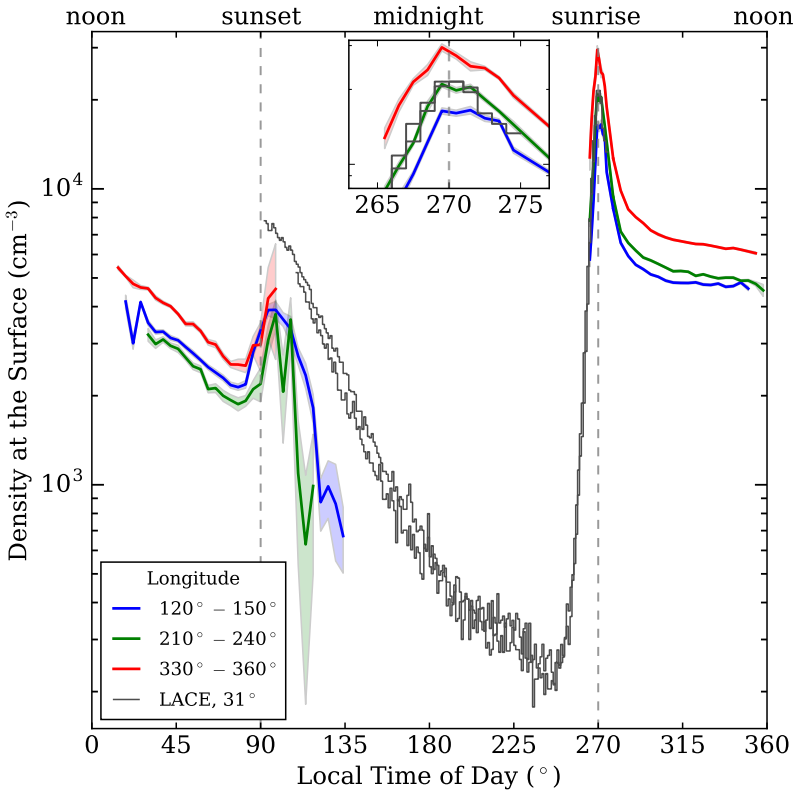


Figure 5.3: The variation of the argon density with local time of day. The coloured lines are typical examples of the LADEE data across the mare (red) and highland (blue and green) selenographic longitude regions, as given in the legend. The shaded areas represent the $\pm 1\sigma$ uncertainties. The inset plot zooms in on the region around the sunrise peak. The grey lines are two sets of LACE data separated by a few lunar days (Hodges, 1975) and normalised to match the density from LADEE at LACE's location at sunrise. A vertical slice would yield the distribution of density with longitude at that time of day, as shown in Fig. 5.4.

times of day is highest at some point over the maria (longitudes from 270°–45°). The peak near sunrise is located over the western maria in the region of the Procellarum KREEP Terrain (PKT), which is rich in ^{40}Ar's parent, ^{40}K (Jolliff et al., 2000). Along the different curves of fixed local time, the selenographic longitude of the peak argon density drifts systematically from ∼300° at sunrise to ∼45° at sunset.

These distributions all reflect a complex interplay between sources, sinks and, most importantly, surface interactions of the argon atoms. Consequently, the variation of argon density with both selenographic longitude and local time of day offer the opportunity to distinguish between different models for the argon exosphere.

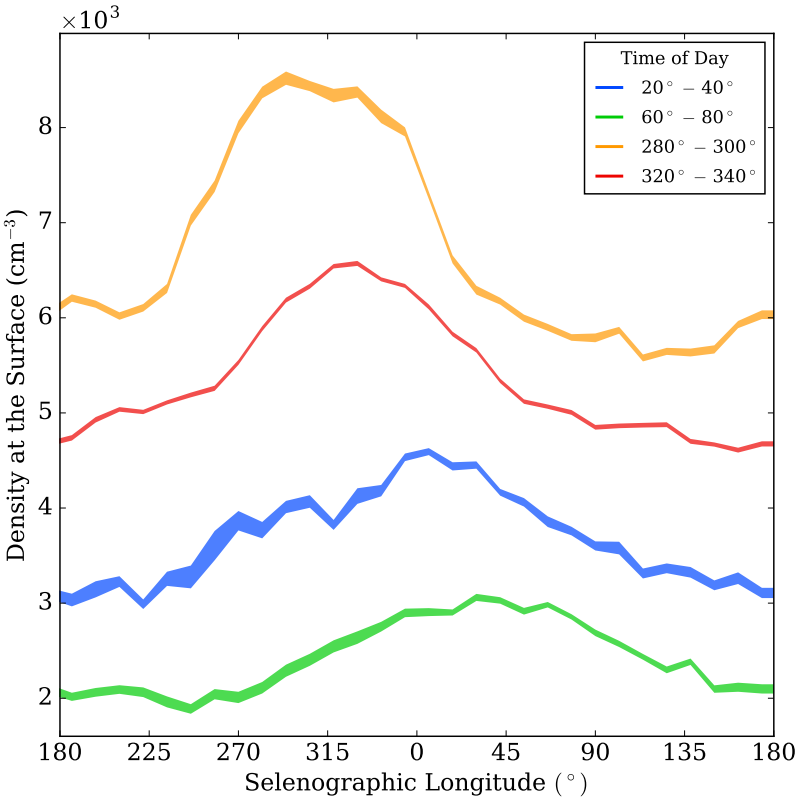


Figure 5.4: The variation of argon density with selenographic longitude from LADEE, showing the bulge over the western maria and its evolution through the lunar day. The coloured regions show a selection of different local times of day, the shaded areas represent the $\pm 1\sigma$ uncertainties. A vertical slice would yield the distribution of density with time of day at that longitude, as shown in Fig. 5.3.

5.3 Model

In this section, we describe the main processes and input parameters in our model. The simulation code itself is publicly available with documentation and examples at icc.dur.ac.uk/index.php?content=Research/Topics/O13.

The central idea is to follow one particle at a time throughout its life, then repeat this for many particles to build up a model of the lunar argon exosphere. Each simulation particle represents a number of argon atoms, and in between its creation and eventual loss from the system, it migrates in a series of interactions with the surface and ballistic hops. The models for each of these various processes are now described in turn.

5.3.1 Source

Most of our simulations are for a steady-state exosphere in which the continuous source rate matches the loss rate, after an initialisation period in which more particles are created than are lost. We assume a continuous source given the relatively smooth variation observed by LADEE, which suggests a lack of dramatic transient source events. The source rate, the mean lifetime, and the total number of particles in the equilibrium system are directly related – knowing any two determines the third. In our simulations, we can investigate a range of mean lifetimes by varying the sinks. The total amount of argon in the exosphere is then scaled to match that from the LADEE measurements by setting the source rate.

The value for the source rate is implicitly varied in a range that reflects the uncertainty in the amount of potassium in the Moon and the effectiveness with which radiogenic ^{40}Ar reaches the surface. Killen (2002) modelled argon's production and diffusion from the potassium in the crust and estimated that argon enters the exosphere at a rate in the range of 3.8–5.5$\times10^{20}$ atoms s^{-1}. These values correspond to only a few per cent of the ^{40}Ar that is created inside the Moon (Hodges, 1975), so we also investigate significantly higher source rates.

Following the discussion from Benna et al. (2015), we wish to test if a (continuous) localised source can reproduce the argon bulge over the western maria and, if so, what source and loss rates this would require. Thus, we use either a global source, where the argon particles appear isotropically at random locations on the spherical surface, or a local source, where they appear preferentially at locations with higher near-surface potassium concentrations. As noted by Benna et al. (2015), while argon is expected to originate from deep, molten sources, there may be preferred diffusion pathways up through the same region marked by the potassium and PKT. The LPGRS, and more recently gamma-ray spectrometers on board Chang'E-1 and Chang'E-2, measured the potassium abundance and distribution in the top metre or so of the regolith (Prettyman et al., 2006; Zhu et al., 2011, 2015). For the localised source model, we use the LPGRS potassium map to weight the source distribution for the simulation particles, such that the probability of being sourced at a given location is proportional to the local potassium concentration.

5.3.2 Sinks

There are two main ways in which particles can be lost from our simulations: interactions with photons or charged particles from the Sun; and cold trapping on the surface in the permanently shadowed polar regions. Our implementation of these

physical processes is described in the following subsections. We include the possibility of gravitational escape in our simulations, but for argon this has a negligible effect.

5.3.2.1 Solar Radiation

A particle in flight on the dayside may be lost due to a variety of processes, the most important of which are photoionisation and charge-exchange with solar wind protons (Grava et al., 2015). An ionised particle will rapidly be driven either away from or into the Moon's surface by the local electromagnetic field. Those that impact the surface may be neutralised and 'recycled' back into the exosphere. Following Butler (1997) and Grava et al. (2015), such processes can be combined to give a single solar radiation destruction timescale, τ. The probability of loss during a flight of time t is then

$$P(t) = 1 - e^{-t/\tau} \, . \tag{5.3.1}$$

For each particle hop, our algorithm picks a random number from a uniform distribution between 0 and 1. If this lies below $P(t)$, then the particle is removed from the simulation.

During LADEE's operation, the mean solar wind speed and proton density were 400 km s^{-1} and 5 cm^{-3} respectively (from the GSFC/SPDF OMNIWeb database interface at omni-web.gsfc.nasa.gov), giving a proton flux of 2×10^8 cm^{-2} s^{-1}. Multiplying this by the interaction cross section, 2×10^{-15} cm^2 (Nakai et al., 1987), gives a rate for proton–argon charge exchange of 4×10^{-7} s^{-1}. Adding the photoionisation rate from Huebner et al. (1992), 3×10^{-7} s^{-1}, taking the inverse, and finally dividing by the recycling fraction of 0.5 (Poppe et al., 2013), gives a timescale of $\tau = 3 \times 10^6$ s. Including the recycling process by simply increasing the timescale is analogous to assuming that a recycled ion reenters the exosphere without travelling a long time or distance. Many of these values have significant uncertainties, such as the cross section which has a somewhat broad peak around 1 keV protons, and the solar wind speed and proton density, which fluctuate significantly. Thus, this error on this timescale is likely at least a factor of 2.

5.3.2.2 Cold Traps

If a particle lands in a permanent cold trap near the poles, then it is assumed to stick there indefinitely. Like previous models, we adopt a stochastic approach. When a particle lands within 15° of the north or south pole, it has a probability of

being trapped, given by the total fractional area of cold traps in that polar region. Various estimates have been made for the size and distribution of cold traps. The appropriate cold trap area depends on the surface interaction of the specific species and the complicated processes that may occur after landing in a cold trap, either to secure or remove the particles (Schorghofer & Aharonson, 2014; Chaufray et al., 2009). As the uncertainty on the effective cold trap area is large, we leave this as a free parameter.

Fig. 5.5 shows the mean lifetime (and corresponding source rate) of argon atoms in the exospheric system for different cold trap areas, keeping the photodestruction timescale constant and using the surface interaction models described in the following subsection. Larger cold traps result in shorter lifetimes, and a larger source rate is then required to maintain the same total number of argon atoms. Also shown in Fig. 5.5, for reference, are some maximum surface temperatures and their corresponding cold trap areas, as inferred from Diviner data (Vasavada et al., 2012). This is done by relating a given cold trap area to the same-size area that never exceeds a certain temperature.

For our default surface interaction model, the total argon content in the simulation is set at 4×10^{28} atoms to match the LADEE abundance measurements. The main uncertainty in this number arises from the long-term and selenographic variations in density measured by LADEE, amounting to about 44%.

5.3.2.3 Seasonal Cold Traps

We include the possibility of seasonal cold traps in our model in addition to the permanent cold traps. These have fractional areas in the north and south polar 15° that grow and shrink periodically as

$$\begin{aligned} f_{\mathrm{N}} &= \max\left[0\,,\; f_{\mathrm{peak}} \sin\left(2\pi\left(t_{\mathrm{yr}} - 0.5\right)\right)\right] \\ f_{\mathrm{S}} &= \max\left[0\,,\; f_{\mathrm{peak}} \sin\left(2\pi\, t_{\mathrm{yr}}\right)\right]\;, \end{aligned} \tag{5.3.2}$$

where f_{peak} is the peak fractional area and t_{yr} is the time in units of years. Thus, the year begins with the vernal equinox and the southern trap reaches its maximum size one quarter of the way through the year, followed by the autumnal equinox and the northern peak in turn. The seasonal traps disappear completely in the summer half of the year for that pole. This asymmetry is required to match the half-year period suggested by the data (see Fig. 5.2). If instead the seasonal traps varied symmetrically in both halves of the year, for example, sinusoidally as in Eqn. 5.3.2 but without truncation, then the summed seasonal trap area of both poles would

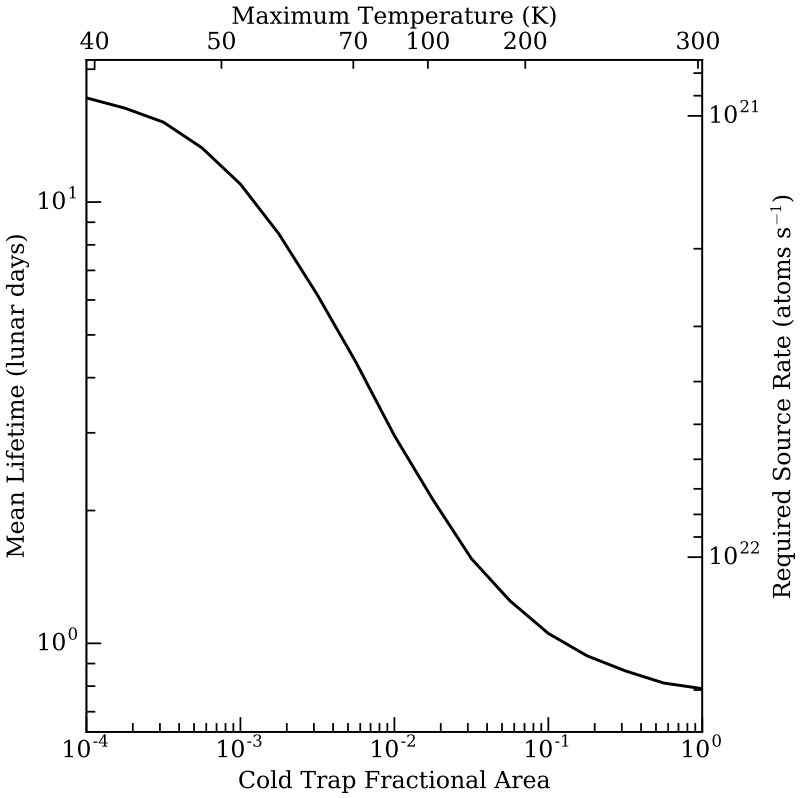


Figure 5.5: The mean lifetime of the simulated argon particles for different cold trap areas as fractions of the polar 15°. The right axis shows the source rate of argon that would be required to match the inferred LADEE total argon abundance. The top axis gives the maximum surface temperatures from Diviner that correspond to the assumed areas covered by cold traps.

vary with a period of 1 year – or be constant if the maximum area at each pole were the same. Thus, a model must have the same overall half-year period to have the potential to explain the data. This requires asymmetrical variation in summer and winter, as is modelled simply by Eqn. 5.3.2.

At the time a particle lands in a seasonal trap, there is a distribution of times at which the present cold traps first appeared (and hence the times at which they will disappear), following Eqn. 5.3.2. Some will have appeared at the beginning of that pole's winter and others potentially just before the particle landed. A particle that lands in a seasonal cold trap is released back into the exosphere at a randomly chosen time, following this distribution.

5.3.3 Hop Trajectory

In the absence of forces other than the gravity from the Moon, the complete paths of particles – including the landing position, time of flight, and position and velocity

at any altitude – can be calculated analytically from the starting location and velocity using Kepler's laws, as derived below in §5.3.3.2. This approach is much less computationally intensive than integrating the paths numerically, even when positions and velocities are also calculated at altitude. This is how we transport particles in our simulation and, apart from the time of flight, it matches the approach of Butler (1997) and Crider & Vondrak (2000) for finding the landing position. For their time of flight calculation, they effectively assumed that the surface is flat, leading to slightly underestimated flight times.

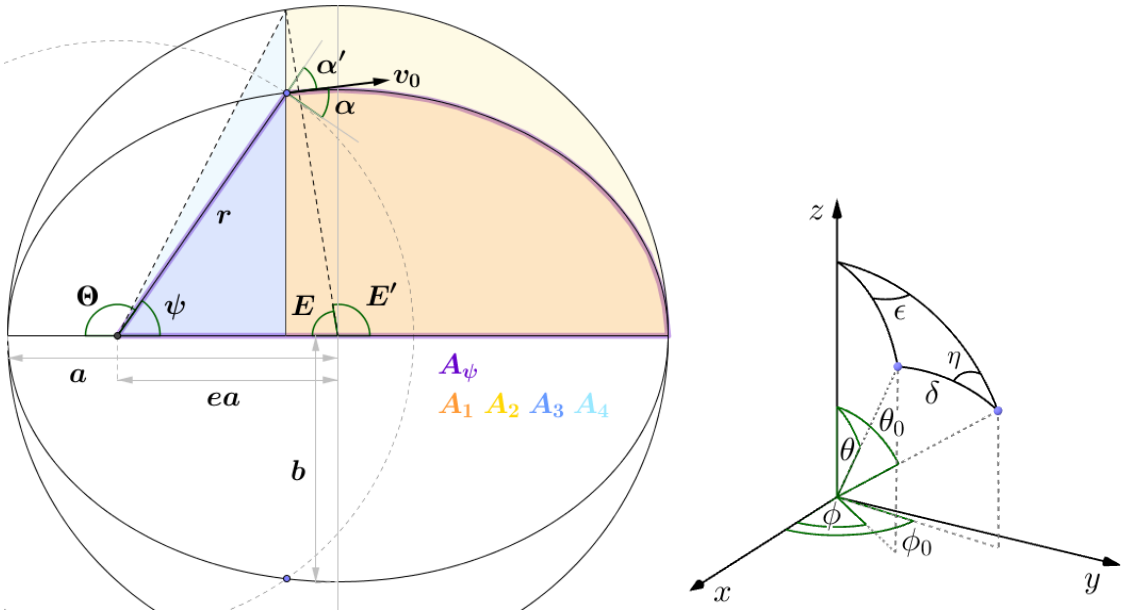
For the initial velocity, the particles are assumed to have accommodated to the surface temperature at their starting location and to have a Maxwell–Boltzmann distribution in the exosphere for that temperature. This means that the initial speed for a particle leaving the surface must be drawn from the Maxwell–Boltzmann flux distribution (Brinkmann, 1970; Smith et al., 1978). As each particle represents a small packet of argon atoms moving through the surface, the random emission direction in the outward hemisphere needs to be weighted by the component of the speed in the vertical direction. The good fit of the simulation densities at altitude to the LADEE data suggests that this model is appropriate. Another difference between our treatment and those of Butler (1997) and Crider & Vondrak (2000) is that they chose emission angles away from vertical, α, from a non-isotropic distribution that was uniform between 0 and $\pi/2$, without including the $\sin(\alpha)$ term that accounts for the full area of the emission hemisphere.

5.3.3.1 Notation

Figs. 5.6a and b show all the relevant notation for a particle's hop. We use standard polar coordinates θ and ϕ, with colatitude $\theta = 0$ at the north pole, and standard selenographic longitude $\phi = 0$ at the sub-Earth point. The local time of day, ϕ', is given by the longitude relative to the subsolar point (noon, where $\phi' = 0$). To account for libration, the longitude of the subsolar point, ϕ_{ss}, is calculated from JPL HORIZONS ephemerides (Giorgini, 2015) (latitude variations are ignored). The time of day is then simply $\phi' = \phi - \phi_{ss}$. If we had ignored libration, then errors of over 3° would have been introduced for some local times of day.

5.3.3.2 Trajectory Derivation

Starting from the initial position and velocity of a particle, we can calculate the landing position (or the position at any altitude) by first finding the parameters of the elliptical path, shown in Fig. 5.6a. The vis-viva equation gives the semimajor



(a) Ellipse and circumcircle notation. (b) Start and end point notation.

Figure 5.6: The notation for the dimensions, angles, and areas for the trajectory calculations: **(a)** a, b, and e are the semimajor axis, semiminor axis, and eccentricity of the ellipse, respectively. The variables Θ and E are the true anomaly and mean anomaly of the object at a distance r from the focus; ψ and E' are their complementary angles. α and v_0 give the velocity of the object; α' is the complementary angle. A_ψ is the area of the outlined triangle + ellipse section; A_1, A_2, A_3, and A_4 are the areas of the same-colour-shaded sections. The Moon is shown by the dashed circle; the blue points mark the start and end of a particle's hop. **(b)** The subscript 0 denotes the initial position. η is the direction west from north of the particle's initial velocity. ϵ is the difference in longitude and δ is the total angle between the start and end points.

axis, a, in terms of the speed, v:

$$v^2 = GM\left(\frac{2}{r} - \frac{1}{a}\right) \quad \Rightarrow \quad a = \left(\frac{2}{r} - \frac{v^2}{GM}\right)^{-1}, \tag{5.3.3}$$

and the ellipticity, e, is found from the velocity angle α':

$$\sin(\alpha') = \frac{r\dot{\psi}}{v} = \sqrt{\frac{a^2(1-e^2)}{2ar - r^2}} \tag{5.3.4}$$

$$\Rightarrow \qquad e = \sqrt{1 - \left(\frac{2ar - r^2}{a^2}\right)\sin^2(\alpha')}\ . \tag{5.3.5}$$

The equation for an elliptical path can then be rearranged to give the angle ψ in terms of a and e:

$$r(\psi) = \frac{a(1-e^2)}{1 - e\,\cos(\psi)} \tag{5.3.6}$$

$$\Rightarrow \qquad \psi = \cos^{-1}\left[\frac{1}{e}\left(1 - \frac{a(1-e^2)}{r}\right)\right] . \tag{5.3.7}$$

In Fig. 5.6b the angles δ and ϵ can be calculated along with the landing coordinates θ and ϕ. At the start of the hop ($\psi = \frac{1}{2}\delta$), Eqn. 5.3.7 becomes

$$\delta = 2\cos^{-1}\left[\frac{1}{e}\left(1 - \frac{a(1-e^2)}{R}\right)\right] , \tag{5.3.8}$$

where R is the radius of the Moon.

The spherical cosine rule gives

$$\theta = \cos^{-1}\left[\cos(\theta_0)\cos(\delta) + \sin(\theta_0)\sin(\delta)\cos(\eta)\right] . \tag{5.3.9}$$

Using the cosine rule again, but with ϵ instead of η, gives

$$\epsilon = \cos^{-1}\left(\frac{\cos(\delta) - \cos(\theta)\cos(\theta_0)}{\sin(\theta)\sin(\theta_0)}\right) . \tag{5.3.10}$$

Then ϕ is simply given by

$$\phi = \begin{cases} \phi_0 + \epsilon & \eta > \pi \\ \phi_0 - \epsilon & \eta < \pi \end{cases} . \tag{5.3.11}$$

The total time of flight is $t = 2t_\psi$, where t_ψ is the time the particle spends tracing out the area A_ψ in Fig. 5.6a with its radial vector. Using general properties of ellipses, it can be shown that $A_\psi \equiv A_1 + A_3 = \frac{b}{a}(A_2 + A_4)$, and therefore

$$A_\psi = \tfrac{1}{2}\,ab\,\left(E' + e\,\sin(E')\right) . \tag{5.3.12}$$

The total period of the orbit, T, is given by Kepler's third law:

$$T = \sqrt{\frac{4\pi^2 a^3}{GM}} . \tag{5.3.13}$$

Kepler's second law states that equal areas are swept out by the radial vector in equal times, therefore,

$$\begin{aligned} t_\psi &= T\,\frac{A_\psi}{A_{\text{ellipse}}} = T\,\frac{\frac{1}{2}ab(E' + e\,\sin(E'))}{\pi ab} \\ &= \frac{T}{2\pi}(E' + e\,\sin(E')) . \end{aligned} \tag{5.3.14}$$

Finally, the eccentric anomaly, E, is related to the true anomaly, Θ, by

$$E = 2\tan^{-1}\left[\sqrt{\frac{1-e}{1+e}}\tan\left(\frac{\Theta}{2}\right)\right] . \tag{5.3.15}$$

$E' = \pi - E$, $\Theta = \pi - \psi$, and here $\psi = \frac{1}{2}\delta$. Therefore,

$$E' = \pi - 2\tan^{-1}\left[\sqrt{\frac{1-e}{1+e}}\tan\left(\frac{\pi - \frac{1}{2}\delta}{2}\right)\right] . \tag{5.3.16}$$

This, with Eqn. 5.3.14, gives the time of flight.

5.3.4 Surface Interaction

The interaction of argon with the surface determines many of the exosphere's characteristics. However, it is a complicated and poorly understood process. In our model, once a particle lands it is assumed to adsorb immediately. It will then reside upon the surface for some amount of time before being released. The residence time and the kinematics of the desorbed particle both depend sensitively upon the temperature at the location of the particle. During the lunar night, the simulated particle may also 'squirrel' down into the regolith, to resurface at some time the following day. This process turns out to be required to match the observed density distribution with local time of day while maintaining realistic desorption energy values, as is discussed in §5.4.1.

The Diviner radiometer on the Lunar Reconnaissance Orbiter (LRO) produced temperature maps of the lunar surface (Vasavada et al., 2012). We use the analytical fit to the Diviner data from Hurley et al. (2015) to make a map of temperature as a function of local time and location in square-degree bins. Following Hurley et al. (2015), we also introduce a longitudinal Gaussian scatter with $\sigma = 4.5°$ into this map to account, statistically, for topographical relief. This represents the temperature effects of, for example, the orientation of slopes near sunrise, which will receive sunlight at different incidence angles, or the positions of ridges or craters that could see sunrise earlier or later respectively.

5.3.4.1 Residence Time and Desorption

Every $\Delta t = 5$ s (a time step is only introduced in this part of the simulation), the temperature is re-calculated for a particle residing on the surface, to account for the Moon's rotation. This allows a residence time, t_{res}, to be found using a standard modified Arrhenius equation (Bernatowicz & Podosek, 1991):

$$t_{\text{res}} = \frac{h}{kT} \exp\left[\frac{Q}{RT}\right] , \tag{5.3.17}$$

where h is Planck's constant, R is the gas constant, T is the temperature, and Q is the desorption energy.

The probability of the particle desorbing from the surface in a given time step is

$$P(t_{\text{res}}) = e^{-\Delta t / t_{\text{res}}} . \tag{5.3.18}$$

If a uniform random number between 0 and 1 exceeds P, then the particle is released and hops again. Otherwise, the simulation time is advanced by Δt and the particle's position and the local temperature are updated. This continues until the particle is released.

For argon, a barrier-free adsorption process is expected, so heats of adsorption and activation energies for desorption can be equated, both corresponding to Q. Experiments with argon on non-lunar aluminosilicates and mineral oxides have shown that the desorption energy is typically around 8–10 kJ mol^{-1} (Matsuhashi et al., 2001). This increases slightly for more Lewis acidic materials, although higher values were obtained for the heat of adsorption, up to 24 kJ mol^{-1} for some acidified mineral oxides, possibly being representative of lower coverages or corresponding to a small fraction of more strongly adsorbing sites. The surface composition of the Moon is dominated by anorthosite, comprised primarily of a variety of silicate minerals (Cheek et al., 2013; Wieczorek et al., 2006). Therefore, these terrestrial experiments provide a reasonable basis for estimating plausible values of Q.

On a low-energy metal oxide surface facet, Dohnálek et al. (2002) calculated the coverage-dependent desorption energy for argon from temperature-programmed desorption (TPD) data and found it to increase from 8 kJ mol^{-1} to around 13 kJ mol^{-1} at very low coverages (where only the highest energy sites should be occupied by argon atoms). If surface diffusion occurs readily and the coverage is low, then these strong adsorption sites may be accessible to all adsorbing argon atoms. Direct calorimetric heats of adsorption on another porous silica yielded Q values of 18 kJ mol^{-1} (again at low surface coverages of argon, Dunne et al., 1996). Interactions with the pristine lunar regolith may be even stronger (Farrell et al., 2015), and Bernatowicz & Podosek (1991) found with a freshly crushed lunar sample that somewhat-higher energies of up to 31 kJ mol^{-1} (7.4 kcal mol^{-1}) are plausible. We conclude that experiments show we should expect argon-regolith interactions to involve energies around 10–30 kJ mol^{-1}. However, until more in-situ experiments are performed, the precise value must be estimated empirically using observations of the exosphere as a whole.

5.3.4.2 Squirrelling

Argon particles enter the exosphere by migrating up through the porous regolith from the Moon's interior (Killen, 2002). We propose that some proportion of the particles residing on the surface will migrate randomly downward as well, to reenter the exosphere at a later time. A somewhat similar process has been discussed in Hodges (1982) and modelled for water ice in polar cold traps by Schorghofer & Aharonson (2014). As is discussed in §5.4.1, we find that such a process is required to explain the LADEE data while maintaining realistic desorption energies.

The persistent reservoir of adsorbed particles residing on the surface at night would therefore act as a source for building up a distribution of particles with depth by the end of the night, which then re-emerge during the day. We assume that particles on the dayside do not adsorb frequently enough for long enough to 'squirrel' in significant numbers.

To check that these timescales arise naturally from the regolith structure and temperature, we estimate the statistical effect this downwards migration could have on the exosphere's distribution with simple order-of-magnitude considerations. The regolith temperature just 2 cm below the surface never drops below 200 K, and by 10 cm only varies by 15 K either side of 255 K (Teodoro et al., 2015).

For particles randomly migrating in the regolith throughout the lunar night ($\sim 10^6$ s), with typical steps of size $\lambda \sim 1$ μm (around and between grains of $\sim\mu$m–mm diameters) and residence times of $t_{\text{res}} \sim 10^{-7}$ s (for $T = 255$ K and $Q = 28$ kJ mol^{-1}; so negligible traversal times of $\sim 10^{-9}$ s at thermal speeds), they will take $N \sim 10^{13}$ steps, random walking a distance of $\lambda\sqrt{N} \sim 1$ m. This implies that a significant number of particles could bury themselves down into the regolith during the night, with the dense source of particles residing on the surface. By symmetry (since the temperatures below the surface are similar at all times), they will take a similar amount of time (the order of half a lunar day) to migrate out of the regolith and reenter the exosphere. This suggests a population of particles that squirrel into the regolith during the night and typically reenter the exosphere during the day.

We use a very simple model to investigate the effect this process has on the exosphere. Any particle adsorbed to the nighttime surface is given a small (constant) probability, P_{sq}, of becoming buried in the regolith. If this happens, then the particle will reenter the exosphere at some time during the following day. That is, the simulation time for the particle is advanced until the Moon has rotated it into a random local time on the dayside, with a uniform probability distribution. The particle 'emerges' residing on the surface, then will desorb and hop as normal.

It is likely that this process is much more complicated in reality, with a dependence on temperature, gradients of temperature and density, and other factors. As one such example, just after sunrise, the subsurface is colder than the surface. This temperature gradient would discourage particles from migrating upward, perhaps delaying the resurfacing of some squirrelled particles until later in the day than would otherwise be expected. Given these uncertain issues, we note the necessary simplicity of this model and use it to explore the plausible effects of the process.

The uncertain surface interaction parameters Q and P_{sq} are varied in the next section, to determine the best values to describe the LADEE and LACE results.

5.4 Results and Discussion

In this section, we compare our simulations with the data introduced in §5.2. First we show how the treatment of surface interactions affects the variation of argon with local time of day. Then we investigate the competing hypotheses to explain the argon bulge over the western maria. In the final part of this section, we study the possibility of seasonal cold traps being the cause of the long-term variation in the LADEE argon abundance.

5.4.1 Distribution with Local Time of Day

The sensitivity of the simulated exosphere to the values of the model parameters describing the surface interactions is shown in Fig. 5.7. In the left panel, the desorption energy, Q, is varied by ∼20% with the squirrelling process switched off, whereas the right panel shows how the distribution changes when the squirrelling probability, P_{sq}, is altered with fixed $Q = 28$ kJ mol^{-1}.

While the higher desorption energy curves best match the nighttime rate of decrease of argon observed by both LACE and LADEE, these model surfaces are so sticky that the timing of the sunrise peak is delayed too far into the day to match the measurements, as highlighted by the inset panel. The sunrise peak position is sensitively dependent upon the model desorption energy. Given that the sunrise peaks are at the same local time over the highlands and maria (§5.2.3), this suggests that the argon interactions with the surface are similar in these different regions.

Fixing the desorption energy at $Q = 28$ kJ mol^{-1} in order to match the sunrise peak position, the model predicts too much exospheric argon late in the night and too little during the day. This provides empirical motivation for the inclusion of the squirrelling process, which allows argon atoms to build up a subsurface population

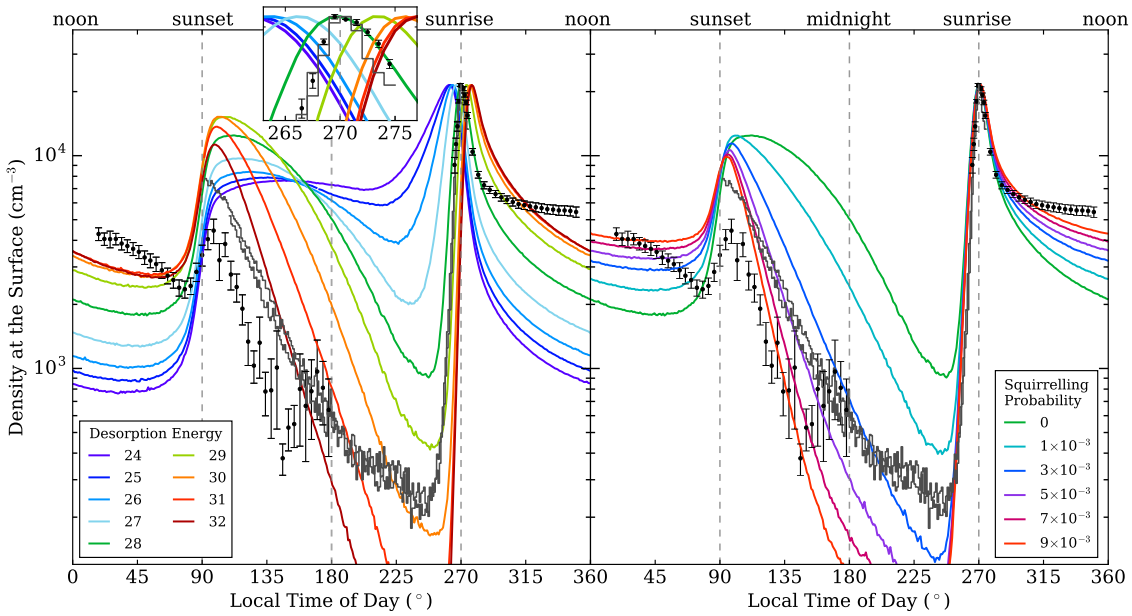


Figure 5.7: The variation of the argon density as a function of local time of day. (left) The desorption energy is varied and no squirrelling is included. The labels show the model desorption energy, Q, in kJ mol^{-1} and the inset plot zooms in on the region around the sunrise peak. (right) $Q = 28$ kJ mol^{-1} and the model squirrelling probability is varied as shown in the legend. In both panels the distributions are normalised to match the peak data density to aid comparison of the shape. The mean data and 1σ errors from LADEE across all longitudes are shown with black points and the LACE data are shown with grey lines.

in the regolith overnight that is released throughout the following day. Increasing $P_{\rm sq}$ enables us to produce a model that matches the nighttime decrease in argon and the shape of the sunrise peak, as shown in Fig. 5.7 (right). For our fiducial value of $P_{\rm sq} = 5 \times 10^{-3}$, the simulation agrees within a factor of 2 with the observations over the entire lunar day. The behaviour from mid-morning to sunset is somewhat discrepant, but given the simplicity of our model and the fact that these details do not change any of the subsequent results and conclusions, we do not complicate the model further.

In contrast to our squirrelling approach, Hodges & Mahaffy (2016) adopted a bespoke temperature-dependent desorption energy (up to ∼120 kJ mol^{-1} at noon) to bring their model into agreement with the LADEE measurements. Introducing all of this freedom into the model can lead to a good fit, but, as Hodges & Mahaffy (2016) themselves noted, such high desorption energies do 'not comport with thermal energies'. The energies required by their model during the day are far beyond the

bond strengths that argon has been measured to make or could be expected to make for the simple van der Waals interactions of a noble gas, as discussed in §5.3.4. Any variable-energy model cannot affect the dayside densities significantly without resorting to these extreme values, because the high dayside temperatures make the residence times negligible for any lower desorption energies. Thus, the nighttime and sunrise densities cannot be simultaneously matched without either including a squirrelling-like process or using unrealistically high, temperature-dependent desorption energies. Note that we also tested a similar model for use in the following bulge and long-term investigations, and the subsequent conclusions were unchanged.

While our squirrelling approach provides a mechanism for fitting the nighttime and daytime argon abundances using physically plausible desorption energies, the simplifications that this model entails should be noted. These processes have not been extensively studied with argon on terrestrial materials, let alone in situ (Dohnálek et al., 2002). In reality, there will be a range of adsorption sites with somewhat different desorption energies, and the probability of adsorption will vary depending upon both the speed of the incoming atom and the presence of volatiles already on the surface. For instance, experiments show that argon has about a 70% probability of adsorbing to a surface at typical exospheric impact speeds of 300 m s^{-1} but much lower temperatures of 22 K (Dohnálek et al., 2002). This probability decreases rapidly with higher impact speeds, reaching zero for 550 m s^{-1}. Argon is also more likely to stick when other argon atoms are already residing on the surface (Head-Gordon et al., 1991). The 300 m s^{-1} adsorption probability reaches one when the coverage is roughly a monolayer. These values may of course be somewhat different for adsorption to the lunar regolith. Note also that the value of Q depends on the exact form of Eqn. 5.3.17. So, it should not be interpreted as a precise estimate of the true energy, especially given the aforementioned details that are all approximated into this single parameter. For example, Grava et al. (2015) used a different prefactor and an extra free parameter, so their fitted value of $Q = 27$ kJ mol^{-1} results in a curve with the sunrise peak around 275°, comparable to our $Q = 30$ kJ mol^{-1}. We focused on fitting the observed sunrise peak time at 270°, so find an effectively lower energy.

We ran additional simulations to investigate the sensitivity of our results to these adsorption issues. Lowering the adsorption probability has a similar effect to lowering the desorption energy. Even for an adsorption probability below 0.1, an increase of a few kJ mol^{-1} to Q results in a similar distribution and sunrise peak position. A speed-dependent adsorption probability also does not dramatically change the distribution, compared with the effects from changing Q or P_{sq}. More importantly, no such mechanisms appear to reduce the need for the squirrelling process to match

the high dayside densities. So, while these known simplifications affect the results at a level that could explain some of the small discrepancies between the model and data distributions, our main conclusions are not sensitive to these assumptions.

One final consideration is the cold trap area and corresponding source rate. The above simulations were all run with the low permanent cold trap fractional area of 0.01% (of the polar 15°) that approximately corresponds to the low source rates estimated by Killen (2002). The much larger traps and high rates considered in the next section have the effect of increasing the density toward the end of the night, which improves the match to the minimum LACE densities (although these may be below LACE's sensitivity (Hoffman et al., 1973)), due to the emergence of newly created particles through the night. The distribution is unaffected at other local times of day.

5.4.2 Longitudinal Bulge

There are two alternative hypotheses for the origin of the argon bulge over the western maria: (1) it reflects a spatially variable source rate that is higher over the western maria (Benna et al., 2015); or (2) it reflects a lower desorption energy from the western maria (Hodges & Mahaffy, 2016). We perform simulations using either a localised source reflecting the LPGRS potassium map (and our $Q = 28$ kJ mol^{-1}, $P_{sq} = 5 \times 10^{-3}$ model described in §5.4.1), or a uniform global source and a spatially varying desorption energy to examine these two scenarios.

For the case of a local source with a greater release of argon over the western maria, the amplitude of the bulge depends sensitively on (1) how localised the source is and (2) the source rate, or – equivalently in the steady state – the loss rate. For a diffuse source, a rapid turnover of particles with short lifetimes is necessary to produce argon atoms that have not travelled so far around the Moon that their locations no longer reflect their origin. We ran simulations with a local source (proportional to the LPGRS potassium abundance) with a range of cold trap fractional areas to determine the area required to produce a bulge comparable with that observed by LADEE.

Fig. 5.8 shows the amplitude of the post-sunrise (280°–300°) argon bulge over the western maria. As shown by the dashed lines, to match the LADEE maximum-to-minimum ratio of 1.5 ± 0.1, a cold trap fractional area of 4 ± 1% of each polar 15° is necessary, corresponding to a mean lifetime of ∼1.4 lunar days. This area is comparable with the 2–2.5% covered by PSRs (Mazarico et al., 2011), but is uncomfortably large if argon is only supposed to be permanently trapped in regions with temperatures never exceeding ∼40 K (Hodges, 1980a). As indicated in Fig. 5.5,

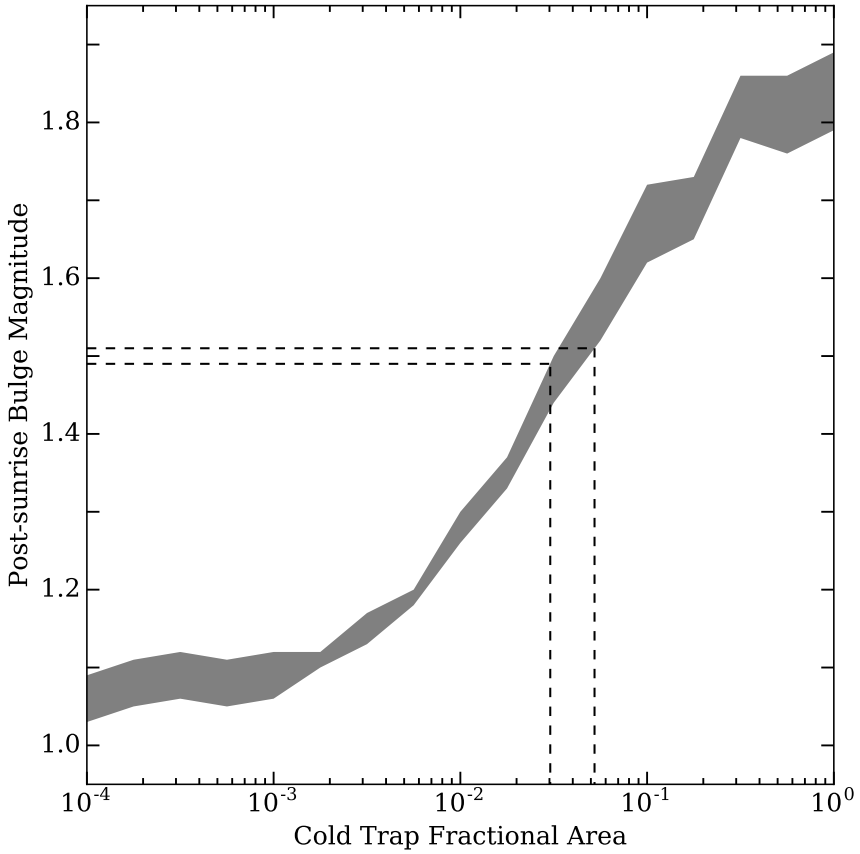


Figure 5.8: The magnitude (the ratio of the maximum to minimum densities over all longitudes at a given time of day) of the simulated bulge over the western maria (in this case at 280°–300° local time of day), generated by the localised source, for different values of the cold trap fractional area of the polar 15°. The dashed lines show the magnitude and uncertainty of the bulge measured by LADEE at the same local time and the corresponding range of cold trap areas.

cold traps of this large size appear to correspond to regions with temperatures that can reach as high as 175 K.

Therefore, for this to be a viable model, one would either need argon to be more readily lost from the exospheric system than previously anticipated, or to have a more highly localised source below the surface than the LPGRS near-surface potassium map. To further investigate the degeneracy between the source rate and how localised the source is, we also tested a 'top-hat' and a point source in the same way. The top-hat source emits argon uniformly from all regions with at least 2,000 ppm of potassium, giving a source region covering 6% of the Moon's surface area in the PKT. This can create an argon bulge with the required amplitude with a cold trap fractional area of only 0.4%, a lifetime of 5.3 lunar days, and a source and loss rate of 2.9×10^{21} atoms s^{-1}. For the extreme case of a point source at 335° longitude

on the equator, a cold trap fractional area of 0.2% is sufficient, with a lifetime of 8.1 lunar days and a rate of 1.9×10^{21} atoms s^{-1}.

If the source is not quite so localised, then feasible causes of increased losses might be: an abundance of small-scale cold traps such as those inferred on Mercury (Paige et al., 2014; McGovern et al., 2013); the presence of other volatiles in PSRs increasing the adsorption probability (Dohnálek et al., 2002); and an unaccounted-for loss mechanism that means that our assumed solar radiation loss rate is an underestimate (recall the large uncertainty on this value as mentioned in §5.3.2). Thus, this model begs an explanation either for the high source and loss rates, or for a highly localised source. Noting this tension, we press on to investigate the shape of the argon bulge in the simulation from the localised (potassium-weighted) source with a cold trap fractional polar area of 4% and compare it with the data.

The variation of argon density with longitude in our simulations is shown in Fig. 5.9 for a sample of different local times of day. The solid lines result from the local source and large cold traps. The dashed lines are from the alternative hypothesis of an isotropic source with mare and highland desorption energies of $Q = 26$ kJ mol^{-1} and 28 kJ mol^{-1} respectively, with a low cold trap fractional area of 0.01% (of the polar 15°) that corresponds approximately to the source rate estimated by Killen (2002). (Note that with an isotropic source the shape of the bulge is insensitive to the source rate.) Also reproduced are the LADEE results. We are predominantly interested in the shapes of the different curves, and not their relative amplitudes, which are determined by the local time of day distribution and are slightly different at certain times, as discussed in §5.4.1. Thus, for clarity, each model curve was divided by its mean and multiplied by the mean of the corresponding data set.

The model with the local source was tuned only to match the maximum-to-minimum ratio of the post-sunrise argon bulge over the western maria observed by LADEE. However, the relative sizes of the bulge at all other times of day, the position, width, and shape of the bulge, and the shift of the bulge to the east throughout the day also happen to be reproduced well. These features result from the fact that the overnight build-up of argon over the western maria around longitudes ∼300°–330° diffuses rapidly across the sunlit surface after it reaches sunrise. This diffusion spreads out the peak in the argon density and shifts it into the dayside: that is, to the east. All these features are also reproduced with the top-hat and point sources, apart from the point source bulge being slightly narrower.

In contrast to these successes of this local source model, the effect of decreasing the adsorption energy over the western maria, shown by the dashed lines in Fig. 5.9, does not match any of the features in the data. The failure to reproduce the observations arises because the lower desorption energy encourages atoms to leave the surface

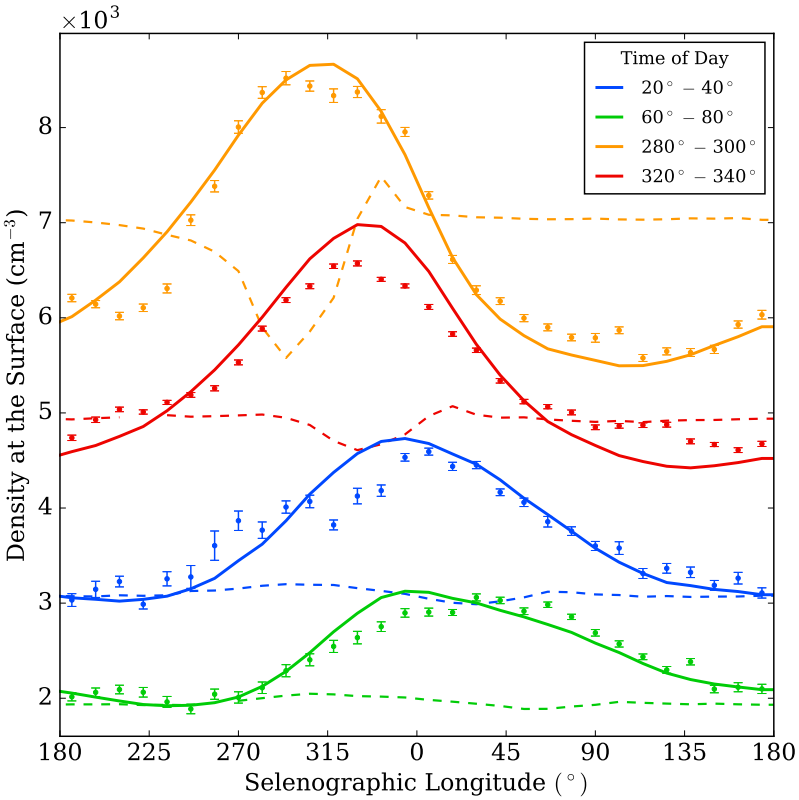


Figure 5.9: The variation of argon density with selenographic longitude for a representative selection of local times of day, shown by the different colours as defined in the legend. The solid lines show the bulge from the local source model with high rates of source and loss, the LADEE data are shown as points, and the dashed lines show the results for the global source with a lower desorption energy in the mare region than in the highlands (26 and 28 kJ mol^{-1} respectively).

and hop more frequently. This is successful in producing a bulge towards the end of the night, where the residence time is longest. However, it also rapidly evacuates the argon from the maria, and the nighttime bulge is replaced by a deficit in argon density almost immediately after sunrise. Therefore, trying to create a local overdensity in this way will inevitably fail if the bulge is required to persist throughout the day, as it is observed to do.

A separate reason to doubt this explanation is that these small changes in adsorption energy lead to significantly different local times for the sunrise peak in the mare and highland regions, in contrast to what the LADEE data show. Consequently, a spatially varying desorption energy explanation for the argon bulge can be ruled out for a couple of reasons. Similar arguments can be used to dismiss the idea that the bulge results from hotter surface temperatures for the maria, for example.

The local source is thus the only proposed hypothesis that has the potential to reproduce the variation of argon density with selenographic longitude seen in the data, and the results are remarkably similar. However, for this explanation to be successful, either (1) the lifetime of an argon atom in the exospheric system (i.e. from source to sink) needs to be ∼1.4 lunar days – if the source rate is proportional to the LPGRS potassium abundance; or (2) the source must be highly localised (or a slightly less extreme combination, as illustrated by the top-hat model). For the diffuse source, the required source rate of 1.1×10^{22} atoms s^{-1} is about 46% of the total rate of argon produced in the Moon (Hodges, 1975), which is unlikely to be able to reach the surface unimpeded. The correspondingly high loss rate that this implies appears to demand widespread polar cold-trapping of argon that exceeds what had previously been considered. Assuming that the cold traps have been stable for ∼1 Gyr (Arnold, 1979), this means that the order of 10^{13} kg of argon would have been delivered to the polar regions during this time.

With this high cold trap fractional area of 4%, it takes roughly six lunar days from the start of the simulation before the exosphere reaches an equilibrium of source and loss rates and a stable number of argon atoms. In comparison, the time it takes for the equilibrium shape of the distributions to become established is always very short, on the order of one lunar day. This timescale is the same even with much smaller cold traps (such as those required for a highly localised source), for which the system can take a long time to reach a true steady state. This is analogous to saying that a localised event diffuses rapidly across the system, even if the total number of particles is still offset from equilibrium.

Irrespective of the lifetime and loss rate of argon in the simulation, particles spend 60% of their life residing on the surface, 30% squirrelling under the surface, and only 10% in flight in the exosphere. Thus, at any given time, these same proportions of the population of argon atoms will be residing on the surface, squirrelling, and in flight. The total number of argon atoms in the simulated exosphere at any time is about 4×10^{28}, corresponding to a mass of 2,600 kg.

5.4.3 Long-Term Variation

Hodges & Mahaffy (2016) suggested that seasonally varying cold traps could produce a periodic signal responsible for the smooth long-term variation in the argon exospheric density by 28% over the 5 months of LADEE's operation. As described in §5.3.2.3, we have included seasonal cold traps to account for the fact that due to the 1.54° obliquity of the Moon, when one pole is tilted away from the Sun, larger areas will act as cold traps for a few months. These seasonal traps would both

temporarily remove and add argon, so that they could drive changes in the density on the relevant timescale without a change in the overall source and loss rates. We performed simulations to test this hypothesis.

The magnitude of this variation is affected by the peak size of the seasonal traps and also by the source and loss rates, which in our model are controlled by the sizes of the permanent cold traps. For the low permanent cold trap fractional area of 0.01% (of the polar 15°) that corresponds approximately to the low source rate estimated by Killen (2002) (where solar wind losses dominate), the 28% change in the density seen by LADEE was reproduced with a peak seasonal cold trap fractional area of $f_{\mathrm{peak}} = 0.8\%$ at each pole's midwinter.

The periodic long-term variation that is produced by these seasonal traps, at the latitudes probed by LADEE, is shown by the solid lines in Fig. 5.10, over a period of 1 year in the simulation. The peak density is predicted to be delayed by about 0.07 years (1 lunar day, ∼1 month) after the minimum trap size at the equinox. This is related to the time it takes for argon to travel between the poles and the equator and is remarkably similar to the delay in the LADEE data after the lunar vernal equinox, as shown in Fig. 5.10. Also shown is how much argon becomes trapped and released at each point in time throughout the year by the seasonal traps. Particles can be trapped from the beginning of winter until the very end, but are only released starting after midwinter, when the traps begin to shrink. This leads to a mild asymmetry in the long-term variation, which could easily be modified by deviating from our simple sinusoidal model.

To demonstrate how the efficacy of the seasonal traps depends on the argon lifetime, we can attempt to reproduce the long-term variation with the very large permanent trap areas (4%) and higher loss rates that, with a localised source weighted by the near-surface potassium, would produce a bulge similar to the LADEE data. We found that much larger seasonal traps would be needed to cause the same magnitude long-term variation, with a peak fractional area of $f_{\mathrm{peak}} = 16\%$. This is because, with much shorter particle lifetimes, fewer particles near the equator have been affected by the seasonal traps. In this case, a maximum of about 2.6×10^{28} argon atoms (over half of the steady-state exosphere) must be temporarily trapped to effect the 28% variation near the equator, over twice as much as in the low loss case. As shown by the dashed line in Fig. 5.10 (top), the delay of the peak density is also affected and occurs 0.04 years (one half lunar day) earlier. This timescale is sensitive to the evolution of the seasonal traps. It is possible that unaccounted-for factors such as thermal inertia, which could cause newly shadowed regions not to begin trapping immediately, delay the variation enough to reproduce the data even with high loss rates.

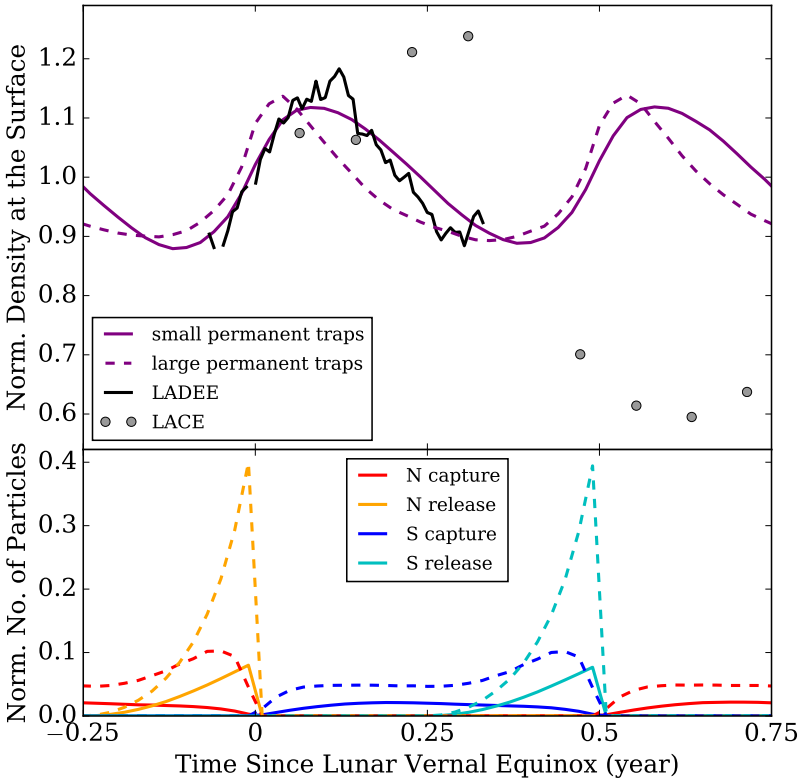


Figure 5.10: (top) The long-term variation of the argon population within 30° latitude of the equator, normalised by the mean density. The purple solid line shows the normalised simulated density at the surface for the low rates model with permanent and peak seasonal cold trap fractional areas of 0.01% and 0.8% respectively; and the dashed line for a high rates simulation with areas of 4% (which reproduced the bulge for the localised potassium-weighted source) and 16%. Time is measured from the last lunar vernal equinox. The LADEE data are shown by the black line for comparison, and the grey points show the magnitudes of the LACE sunrise peak densities (Hodges, 1975) relative to the mean sunrise density measured by LADEE. (bottom) The number of argon particles that are being captured and released in each simulation by the northern and southern seasonal traps, given as a fraction of the total number of argon atoms in the equilibrium system (4×10^{28}). The dashed lines show the results for the large permanent traps.

Unfortunately, the success of this seasonal model with the LADEE data does not continue for the long-term variation measured by LACE, which is shown by the grey points in Fig. 5.10 (assuming it is not the result of instrument degradation). The seasonal variation acts in the opposite way to the trend seen in 1973. Therefore, if

there are seasonal cold traps that explain the LADEE argon data and they were active during the LACE measurements, then the loss rate required to match the drop measured by LACE needs to be significantly larger than had been anticipated by Grava et al. (2015). It is also noteworthy that the later LACE measurements fall well below the minimum measured by LADEE at the same location. This suggests that, if the LADEE data do show a periodic feature and there was not a transient loss event to drive down the later LACE measurements, then the equilibrium size of the exosphere must have increased over the last 40 years.

It is also possible that the LADEE and LACE long-term variations are both the result of transient source events, as suggested by Benna et al. (2015) and Grava et al. (2015). If we assume that the minima of the LADEE and LACE measurements indicate the equilibrium states throughout the measurement periods, and that a transient source had increased the density to the maximum before discontinuing, then we can model the decrease as a simple exponential decay from the maximum measurement back to the equilibrium state. In both the LADEE and LACE cases this would require a lifetime of around 0.9 lunar days, even shorter than that required for our local source bulge hypothesis and implying even greater loss rates. If the equilibrium level is lower than the minimum measurements, then the variations would therefore be part of even larger but less rapid declines. In the absolute limit of no background exosphere at all, the required lifetimes could extend up to nine and five lunar days for LADEE and LACE respectively. In this extreme case, these long lifetimes would require only small source and loss rates and correspondingly small permanent cold traps with temperatures below 70 K, as shown by Fig. 5.5. Similar arguments can be made regarding transient loss events, since LADEE shows an equally rapid increase of argon. Of course, it is imaginable that a combination of multiple, dramatic source and loss events could produce these variations regardless of the lifetime, but this is extraordinarily unlikely.

As discussed in §5.3.2.3, asymmetrical variation of the seasonal traps in summer and winter is necessary to match the half-year period suggested by the data. Therefore, our model included no seasonal cold traps throughout the summer half of the year at each pole. Had we instead allowed some seasonal traps to decrease until the summer solstice, their reduction would have offset some of the effect of the growing traps at the other pole. Consequently, the model would have needed larger seasonal trap areas to match the amplitude of the observed long-term argon variation. It should be possible to use lunar elevation maps to predict the actual variation of seasonal cold traps throughout the year, at least enough to test whether such an asymmetrical variation is realistic. In the scope of this work, we show only that this hypothesis has the potential to explain the data.

5.5 Conclusions

In this chapter, we have studied the LADEE (and LACE) measurements of the lunar argon exosphere and developed a Monte Carlo model to investigate what is implied about the sources, sinks, and surface interactions in the system, and to test whether various hypotheses are able to explain the observed features.

The extrapolation of simulated density at an altitude of 60 km to density at the surface is fitted to within 12% everywhere using a model of two Chamberlain distributions with different temperatures. From this altitude, using a single Chamberlain distribution or its first order approximation can lead to overestimates greater than a factor of 3. These errors can be much larger for extrapolations from higher altitudes. Other exospheric species should exhibit similar behaviour. Lighter particles typically travel farther each hop, which would increase the error from using a single-temperature model. The two-temperature model fits the LADEE data well, suggesting that simple thermal desorption dominates the release energetics of exospheric argon.

The distribution with local time of day of the argon density in the exosphere is very sensitive to the nature of the interactions with the surface. Apart from an offset in amplitude reflecting the higher density over the maria, the highland and mare results are very similar, suggesting that the surface interactions do not differ greatly with regolith composition at equatorial latitudes. A very simple model allowing atoms to squirrel into the regolith overnight, building up a subsurface population that is released during the following day, can reproduce the broad characteristics of the observed exosphere at all times of day, without the need to resort to unreasonably high and temperature-dependent desorption energies. The timing of the sunrise peak requires a residence time near sunrise of 1,300 s, which corresponds to a desorption energy of 28 kJ mol^{-1}, a high but plausible value for noble gas interactions. The subsequent results are insensitive to the details of these surface interaction models.

Of the two hypotheses that have been proposed for the origin of the argon bulge over the western maria and PKT, only a localised source has the potential to explain this feature. Our simulations with this model can reproduce the observed size, shape, and position of the bulge at all local times of day. There is a degeneracy between how localised to the mare region the source is and the lifetime and rates that the data require. For a source distribution weighted by the LPGRS potassium map, the observed bulge is reproduced with a mean lifetime for argon atoms in the exospheric system of only 1.4 lunar days, corresponding to a high equilibrium source and loss rate of 1.1×10^{22} atoms s^{-1}. To achieve this, our model would need permanent cold traps that have a total area comparable with the PSRs measured at Diviner's

resolution, or some other additional loss mechanism. A more highly localised source can reduce the required rates and trap areas by an order of magnitude – a point source reproduces a bulge of the right amplitude with a source and loss rate of 1.9×10^{21} atoms s^{-1}. So, despite this model's unique success in reproducing the data, it begs an explanation for some combination of source localisation and high source and loss rates.

Models that aim to create the argon bulge by encouraging atoms to hop either more frequently or higher founder because they naturally lead to a short-lived feature through the night that is replaced by a local deficit in the argon density after sunrise.

The long-term variation in the global argon density seen by LADEE can be elegantly explained by the periodic behaviour of seasonal cold traps. The details of how large they need to be depend upon the base source and loss rates. The time lag of the peak density in the data is reproduced naturally by the model for small cold traps and low rates. It is slightly offset for higher rates, which might be mitigated by effects such as thermal inertia. However, the long-term decrease seen by LACE in 1973, if real, requires some other significant source and/or loss process because the seasonal variation should act in the opposite way to the observed trend. The relatively smooth variation of the argon density observed over the lifetime of LADEE suggests that significant transient release or loss events are unlikely to be the cause. This includes the apparent lack of a significant effect from the periodic crossing of the Moon through the Earth's magnetotail, which might have been expected to reduce the solar wind loss rate. Any transient source (or loss) explanation would also require high rates of source and loss for the system to return to equilibrium after the event within the measured lunar-day timescales, unless the equilibrium density is far lower than the minimum observed by LADEE.

Seasonal cold traps should be expected to impact other species in the exosphere in a similar way, depending on their threshold trapping temperature. If any non-radiogenic, condensible species (such as methane) (Hodges, 2016) were also found to follow the variation seen for argon, then this would be strong evidence in support of the seasonal hypothesis (and vice versa). This is because tidal or seismic changes that might affect the argon source rate would be irrelevant for species that do not come from inside the Moon. Further long-term observations of the argon density would also help determine whether the variation is actually periodic in the first place.

There are several experiments that could help determine what combination of source localisation and rate of source and loss is responsible for the bulge, given the lack of other possible explanations. To test the hypothesis of a diffuse localised source with very high source and loss rates, one could pursue: in-situ searches for argon

trapped in PSRs, although there are various uncertainties regarding how much sequestered argon would be found and at what depth (Schorghofer & Taylor, 2007); or measurements of the exosphere towards the poles, where the effects that large cold traps would have on the distribution of argon with latitude would be detectable. On the other end of the degeneracy: if the source is highly localised, then the large differences in that rate should directly affect the late-night argon distributions in the mare and highland regions. This might need to be measured at or near the surface to detect the very low densities. Future investigations of this kind would help determine if this model is indeed the origin of the bulge, or if some entirely new explanation is required.

Chapter 6

Conclusions

6.1 Summary and Context

The Solar System used to be a much more violent place. The importance of the giant impacts that dominate the final stages of planet formation has been appreciated for several decades. However, we still have a limited understanding of the detailed consequences of any specific collision.

In this thesis, I have tackled both some of these specific unanswered questions about giant impacts and some of the methodological challenges that lurk behind the scenes. The primary aims were to explore the ramifications of the collision that knocked the planet Uranus over to spin on its side; to study the numerical convergence of model results with an unprecedented number of simulation particles; and to explore for the first time the detailed consequences of giant impacts onto planets with thin atmospheres.

In Chapter 2, I described the development of the required tools: SWIFT, a smoothed particle hydrodynamics (SPH) code capable of simulating more than 1000× the number of particles used in standard simulations of giant impacts in the literature; and SEA, a new method to quickly create relaxed spherical arrangements of particles for initial conditions. Both of these codes have been publicly released.

In the first half of Chapter 3, I focused on the earliest planetary oddity blamed on a giant impact: Uranus' uniquely high obliquity. With a suite of SPH simulations of collisions with different mass impactors and different angular momenta, I confirmed that a wide variety of giant impacts can readily produce the spin we see today. Furthermore, a subset of grazing impacts might help explain the unsolved mystery of Uranus' cold exterior, by trapping the interior heat below a thermal boundary layer. I also found that the majority of the atmosphere can survive the impact, with the first 3D simulations of atmospheric erosion from more than just head-on collisions.

At the time that this work was published, it was the only project to have run such simulations since the single previous study in 1992. In the two short years between then and the time of writing this thesis, two other groups have submitted their own complementary impact simulations on the topic (Kurosaki & Inutsuka, 2019; Reinhardt et al., 2019), and multiple mission proposals to visit both Uranus and Neptune have been gaining support (e.g. Fletcher et al., 2019). It has become an exciting time to be studying the ice giants. Not only are they the only planets in our solar system to have never had a dedicated mission, but they represent prime examples of what appears to be the most common class of exoplanets (Batalha, 2014).

At this point, making use of the code developments presented in Chapter 2, I improved the mass resolution of the simulations by a factor of 1000. In the second half of Chapter 3, I demonstrated that standard-resolution SPH simulations with 10^5–10^6 particles can fail to produce reliable answers to questions about even the large-scale outcomes of giant impacts. Simulations with 10^7 and just over 10^8 particles demonstrated apparently converged results for all the major processes in collisions on Uranus. However, small-scale features such as the precise composition of ejecta can require even higher resolutions for convergence.

These results align with those of Hosono et al. (2017), whose study is the only other to date to run simulations using particle numbers in these orders of magnitude. Consequently, this research opens as many questions as it answers. Chief among which is: How many previous conclusions regarding the consequences of giant impacts are actually reliable? And how many particles do we need to adequately resolve different phenomena? Hopefully, in light of these recent developments, convergence tests will become a routine requirement for impact studies. Next-generation tools like SWIFT should make this an achievable goal even for topics requiring high resolutions.

Moving on from these important but perhaps less inspiring tests of old results, in Chapter 4 I presented some new steps towards answering the many open questions about atmospheric erosion by giant impacts. Previous studies have primarily been limited to 1D models. With the high particle numbers made possible by SWIFT, this suite of simulations explored the wide parameter space of impacts onto an atmosphere-hosting proto-Earth with different speeds and impact angles. These violent collisions are complicated, and the simulations reveal a variety of important and distinct processes by which the target planet can lose atmosphere, depending on the specific scenario.

I found that there is a gradual transition from removing almost none to almost all of the atmosphere for a grazing impact as it becomes more head-on or increases

in speed, including complex, non-monotonic behaviour at low impact angles. In contrast, for head-on impacts, a slightly greater speed can suddenly remove much more atmosphere. The quantitative results agree quite well with 1D estimates of local erosion applied to the ground speeds measured in the simulation. Predictions of the global loss from an idealised shock wave come close to matching the cleanest, highly grazing impacts, but cannot account for the dramatic disruption of the target in more head-on collisions, nor the fate of material that the impactor encounters directly. This interplay of multiple non-linear and asymmetrical loss mechanisms highlights the need for 3D simulations in order to make realistic predictions.

This project is an initial step towards improving our understanding both of the evolution of Earth and its atmosphere following the Moon-forming collision, and of the role that giant impacts might play in creating the striking diversity of atmospheres hosted by exoplanets.

Finally, in Chapter 5, I presented the first tests of competing explanations for the longitudinal and long-term variations in the lunar argon exosphere that were discovered by the recent LADEE mission. I found that the proposed spatial variations of argon–surface interactions fail to reproduce the observations, and that the persistent argon enhancement over the western maria can only be explained by a localised source. Depending on the extent of the source, it might also need to produce argon at an extremely high rate. Either characteristic would have intriguing implications for the distribution – and therefore origin – of the argon's potassium parent inside the Moon. Independently, the unexpected long-term changes of the global argon density could be explained by seasonal fluctuations of the polar cold traps, but more data are needed to confirm whether the behaviour is genuinely periodic.

Surface-bounded exospheres like this one are also common throughout the Solar System. So these revelations about its complex evolution on both short and long timescales help to shape the modelling and understanding of many other systems. One notable example is the Moon's water exosphere, which contains much less mass but follows the same processes of migration and loss (Benna et al., 2018). While the more abundant argon is much easier to observe, water, especially if accumulated in polar craters, would be an invaluable resource for future lunar bases.

6.2 Future Work

Most targets for future research fit neatly into three categories: improving existing approaches; coupling to other techniques; and pursuing new topics.

In the context of the projects in this thesis, the immediate inclination in the first category is to run SPH simulations using ever greater numbers of particles. With more codes like SWIFT being developed to take advantage of contemporary supercomputing architectures for specific applications while also being made fully public, this will hopefully be the ongoing reality. However, as discussed in Chapter 2, simulations of giant impacts have several significant shortcomings that can and should be improved upon. Sophisticated modifications to the basic formulation of SPH have been proposed and tested in planetary science and for many other applications, often attempting to solve similar problems such as the mixing of different materials or the treatment of free surfaces (e.g. Hosono et al., 2016; Deng et al., 2019). Material strength has also been successfully included in SPH (e.g. Benz & Asphaug, 1999), and far more realistic equations of state are being developed alongside improving experimental results and theoretical understanding of molecular dynamics (e.g. Stewart et al., 2008). Finally, more careful comparisons with other methods such as grid-based schemes will help both to identify weaknesses and to promote the most appropriate tools for different computational jobs.

The Uranus simulations in Chapter 3 are a perfect starting point for the second category of coupling to other techniques. SPH allows us to predict the immediate aftermath of an impact. To draw meaningful conclusions about what happened to any specific planet, these results must be connected to present-day observations. In the case of Uranus' absent flow of heat from the interior and the theoretical hypothesis of a thermal boundary layer, the post-impact output from SPH simulations could be used as the input for models of long-term thermal evolution. For the origin of Uranus' satellites, the evolution of the post-impact debris disk could be modelled in the same way as has been done to study the accretion of Earth's moon (e.g. Pahlevan & Stevenson, 2007). In addition, the fate of any pre-existing satellites could be studied by simulating their evolution in the gravitational influence of the post-impact planet, including any disruption or drag caused by debris they might encounter. This is also an opportunity to make the most of the high-resolution simulations, with far greater detail available for the structure, composition, and thermal state of the debris than was previously possible.

Many similar opportunities exist for the atmospheric erosion simulations of Chapter 5. The thermal state of the post-impact planet can be used to estimate the subsequent heating and atmospheric loss by hydrodynamic escape, as well as the geophysical evolution of the mantle (e.g. Abe & Matsui, 1986). In the context of exoplanets, these simulations could also help to predict the observability of giant impacts. The bursts of energy from collisional heating and the significant increase in the planet's effective size might be directly detectable (Dokuchaev & Eroshenko, 2013). Furthermore, the

debris that continues to orbit around the star could be observable for a long time after the impact itself (Wyatt & Jackson, 2016). Comparing these predictions for observability with detection rates from new telescopes like WFIRST and E-ELT (Lacy et al., 2018; Udry et al., 2014) would constrain models of planet formation by the frequency of giant impacts in other systems. SPH predictions for the outcomes of different impacts can also be used to improve simulations of planetary accretion, many of which currently assume that colliding proto-planets perfectly stick to each other (e.g. Chambers, 2013).

Finally, in the third category, we can apply the same techniques to answer different questions. For topics like atmospheric erosion, the field is new enough that extending simulations to any other region of the huge parameter space would be valuable. For example: more massive targets like super-Earths and mini-Neptunes; different-mass impactors that may host their own atmospheres; or different atmospheric compositions. Adding a primordial ocean would allow tests of the predicted increase in lost atmosphere and provide constraints on the Earth's volatile history (Genda & Abe, 2005). Or one could return to the long list of suspected giant impacts throughout the Solar System, from lunar formation and Mars' crustal dichotomy to Saturn's mid-sized moons and the Pluto–Charon system (e.g. Wyatt & Jackson, 2016).

The link between Chapter 5's comparatively standalone exosphere project and giant impacts is another opportunity for future progress. I placed constraints on several aspects of the argon source in the Moon's interior, but with further modelling the degeneracy might be broken between how localised and how rapid this source is, especially if combined with data from new missions. This provides a novel way to investigate the argon-producing KREEP and its origin, presumably from the hot magma ocean that was itself a result of the giant impact that created the debris disk (Wieczorek et al., 2006). This is not likely to be the most immediate source of new information about the Moon-forming collision, but it does serve as an extra background incentive for studying our closest example of surface-bounded exospheres.

In conclusion, it is an exciting time to be involved in studying planet formation. Simulations are rapidly becoming more sophisticated, allowing us to perform ever more realistic 'experiments', and observations are revealing an enormous number and diversity of exoplanets. We are no longer restricted to the study of a handful of middle-aged planets in our solar system, hunting for evidence of their violent early lives. This thesis has presented some examples of using numerical simulations to learn a little more about the consequences of giant impacts for an ice giant and for rocky planets with thin atmospheres. There is a lot more to do.

Bibliography

Abe Y., Matsui T., 1986, J. Geophys. Res., 91, E291

Agertz O., et al., 2007, Mon. Not. R. Astron. Soc., 380, 963

Ahrens T. J., 1993, Annu. Rev. Earth Planet. Sci., 21, 525

Archinal B. A., et al., 2011, Celest. Mech. Dyn. Astron., 109, 101

Arnold J. R., 1979, J. Geophys. Res., 84, 5659

Asphaug E., 1997, Meteorit. Planet. Sci, 32, 965

Asphaug E., Agnor C. B., Williams Q., 2006, Nature, 439, 155

Balsara D. S., 1995, J. Comput. Phys., 121, 357

Batalha N. M., 2014, Proc. Nat. Aca. Sci. USA, 111, 12647

Benna M., Mahaffy P. R., Halekas J. S., Elphic R. C., Delory G. T., 2015, Geophys. Res. Lett., 42, 3723

Benna M., Elphic R., Stubbs T., Hurley D., Mahaffy P., 2018, in 42nd COSPAR Scientific Assembly. pp B3.1–8–18

Benz W., Asphaug E., 1999, Icarus, 142, 5

Benz W., Slattery W. L., Cameron A. G. W., 1986, Icarus, 66, 515

Benz W., Anic A., Horner J., Whitby J. A., 2007, Space Sci. Rev., 132, 189

Bernatowicz T. J., Podosek F. A., 1991, in LPSC Proceedings. pp 307–313

Biersteker J. B., Schlichting H. E., 2019, Mon. Not. R. Astron. Soc., 485, 4454

Borrow J., Bower R. G., Draper P. W., Gonnet P., Schaller M., 2018, Proc. 13th SPHERIC Intl. Wksh., pp 44–51

Brinkmann R. T., 1970, Planet. Space Sci., 18, 449

Butler B. J., 1997, J. Geophys. Res., 102, 19283

Cameron A. G. W., 1985, Icarus, 62, 319

Cameron A. G. W., Ward W. R., 1976, in LPSC. p. 120

Canup R. M., Ward W. R., 2006, Nature, 441, 834

Canup R., Barr A. C., Crawford D. A., 2013, Icarus, 222, 200

Chamberlain J. W., 1963, Planet. Space Sci., 11, 901

Chambers J. E., 2001, Icarus, 152, 205

Chambers J. E., 2013, Icarus, 224, 43

Chaufray J.-Y., Retherford K. D., Gladstone G. R., Hurley D. M., Hodges R. R., 2009, in Lunar Reconnaissance Orbiter Science Targeting Meeting. pp 22–23

Cheek L. C., Donaldson Hanna K. L., Pieters C. M., Head J. W., Whitten J. L., 2013, J. Geophys. Res. (Planets), 118, 1805

Citron R. I., Genda H., Ida S., 2015, Icarus, 252, 334

Collins G. S., 2014, J. Geophys. Res. (Planets), 119, 2600

Crider D. H., Vondrak R. R., 2000, J. Geophys. Res., 105, 26773

Ćuk M., Stewart S. T., 2012, Science, 338, 1047

Cullen L., Dehnen W., 2010, Mon. Not. R. Astron. Soc., 408, 669

Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, Astrophys. J., 292, 371

Dehnen W., Aly H., 2012, Mon. Not. R. Astron. Soc., 425, 1068

Deng H., Reinhardt C., Benitez F., Mayer L., Stadel J., Barr A. C., 2019, Astrophys. J., 870, 127

Diehl S., Rockefeller G., Fryer C. L., Riethmiller D., Statler T. S., 2015, Publ. Astron. Soc. Aust, 32, e048

Dohnálek Z., Scott Smith R., Kay B. D., 2002, J Phys. Chem. B, 106, 8360

Dokuchaev V. I., Eroshenko Y. N., 2013, Planet. Space Sci., 78, 64

Dunne J. A., Mariwala R., Rao M., Sircar S., Gorte R. J., Myers A. L., 1996, Langmuir, 12, 5888

Elphic R. C., Delory G. T., Hine B. P., Mahaffy P. R., Horanyi M., Colaprete A., Benna M., Noble S. K., 2014, Space Sci. Rev., 185, 3

Farrell W. M., et al., 2015, in Space Weathering of Airless Bodies: An Integration of Remote Sensing Data, Laboratory Experiments and Sample Analysis Workshop. p. 2037

Fletcher L. N., et al., 2019, arXiv e-prints, p. arXiv:1907.02963

Fortney J. J., Mordasini C., Nettelmann N., Kempton E. M.-R., Greene T. P., Zahnle K., 2013, Astrophys. J., 775, 80

Fressin F., et al., 2013, Astrophys. J., 766, 81

Fryer C. L., Rockefeller G., Warren M. S., 2006, Astrophys. J., 643, 292

Genda H., Abe Y., 2003, Icarus, 164, 149

Genda H., Abe Y., 2005, Nature, 433, 842

Genda H., Fujita T., Kobayashi H., Tanaka H., Abe Y., 2015, Icarus, 262, 58

Gingold R. A., Monaghan J. J., 1977, Mon. Not. R. Astron. Soc., 181, 375

Giorgini J. D., 2015, IAU General Assembly, 22, 2256293

Gomes R., Levison H. F., Tsiganis K., Morbidelli A., 2005, Nature, 435, 466

Gonnet P., 2015, SIAM J Sci. Comput., 37, C95

Grava C., et al., 2015, Icarus, 255, 135

Halliday A. N., 2013, Geochem. Cosmo. Acta, 105, 146

Hartmann W. K., Davis D. R., 1975, Icarus, 24, 504

Head-Gordon M., Tully J. C., Schlichting H., Menzel D., 1991, J. Chem. Phys., 95, 9266

Herant M., 1994, Mem. Soc. Astron. It., 65, 1013

Hodges Jr. R. R., 1975, Moon, 14, 139

Hodges Jr. R. R., 1980a, in Bedini S. A., ed., LPSC Proceedings Vol. 11, LPSC Proceedings. pp 2463–2477

Hodges R. R., 1980b, J. Geophys. Res., 85, 164

Hodges R. R., 1982, in LPSC Proceedings. pp 329–330

Hodges R. R., 2016, Geophys. Res. Lett., 43, 6742

Hodges Jr. R. R., Johnson F. S., 1968, J. Geophys. Res., 73, 7307

Hodges R. R., Mahaffy P. R., 2016, Geophys. Res. Lett., 43, 22

Hoffman J. H., Hodges Jr. R. R., Johnson F. S., Evans D. E., 1973, in LPSC Proceedings. p. 2865

Hosono N., Saitoh T. R., Makino J., Genda H., Ida S., 2016, Icarus, 271, 131

Hosono N., Iwasawa M., Tanikawa A., Nitadori K., Muranushi T., Makino J., 2017, Publ. Astron. Soc. Jpn., 69, 26

Hubbard W. B., MacFarlane J. J., 1980, J. Geophys. Res., 85, 225

Hubbard W. B., Marley M. S., 1989, Icarus, 78, 102

Huebner W. F., Keady J. J., Lyon S. P., 1992, Astrophys. Space Sci., 195, 1

Hurley D. M., Sarantos M., Grava C., Williams J.-P., Retherford K. D., Siegler M., Greenhagen B., Paige D., 2015, Icarus, 255, 159

Hurley D. M., et al., 2016, Icarus, 273, 45

Hwang J., Chatterjee S., Lombardi James J., Steffen J. H., Rasio F., 2018, Astrophys. J., 852, 41

Inamdar N. K., Schlichting H. E., 2015, Mon. Not. R. Astron. Soc., 448, 1751

Inamdar N. K., Schlichting H. E., 2016, Astrophys. J., 817, L13

Jolliff B. L., Gillis J. J., Haskin L. A., Korotev R. L., Wieczorek M. A., 2000, J. Geophys. Res., 105, 4197

Kegerreis J. A., Eke V. R., Massey R. J., Beaumont S. K., Elphic R. C., Teodoro L. F., 2017, J. Geophys. Res. (Planets), 122, 2163

Kegerreis J. A., et al., 2018, Astrophys. J., 861, 52

Kegerreis J. A., Eke V. R., Gonnet P., Korycansky D. G., Massey R. J., Schaller M., Teodoro L. F. A., 2019, Mon. Not. R. Astron. Soc., 487, 1536

Kenworthy M. A., Mamajek E. E., 2015, Astrophys. J., 800, 126

Killen R. M., 2002, Meteorit. Planet. Sci, 37, 1223

Kipping D. M., Nesvorný D., Buchhave L. A., Hartman J., Bakos G. Á., Schmitt A. R., 2014, Astrophys. J., 784, 28

Korycansky D. G., Bodenheimer P., Cassen P., Pollack J. B., 1990, Icarus, 84, 528

Kraus R. G., Root S., Lemke R. W., Stewart S. T., Jacobsen S. B., Mattsson T. R., 2015, Nature Geoscience, 8, 269

Kurosaki K., Inutsuka S.-i., 2019, Astrophys. J., 157, 13

Lacy B., Shlivko D., Burrows A., 2018, preprint, p. arXiv:1801.08964

(arXiv:1801.08964)

Lawrence D. J., Feldman W. C., Barraclough B. L., Binder A. B., Elphic R. C., Maurice S., Thomsen D. R., 1998, Science, 281, 1484

Lecoanet D., et al., 2016, Mon. Not. R. Astron. Soc., 455, 4274

Leopardi P., 2007, PhD thesis, Sch. Math. Stat., U. New South Wales

Levison H. F., Morbidelli A., Tsiganis K., Nesvorný D., Gomes R., 2011, Astrophys. J., 142, 152

Liu S.-F., Hori Y., Lin D. N. C., Asphaug E., 2015, Astrophys. J., 812, 164

Lock S. J., Stewart S. T., Petaev M. I., Leinhardt Z., Mace M. T., Jacobsen S. B., Cuk M., 2018, J. Geophys. Res. (Planets), 123, 910

Lombardi J. C., Sills A., Rasio F. A., Shapiro S. L., 1999, J. Comput. Phys., 152, 687

Lopez E. D., Fortney J. J., 2014, Astrophys. J., 792, 1

Lopez E. D., Fortney J. J., Miller N., 2012, Astrophys. J., 761, 59

Lucy L. B., 1977, Astrophys. J., 82, 1013

Lupu R. E., et al., 2014, Astrophys. J., 784, 27

Mahaffy P. R., et al., 2014, Space Sci. Rev., 185, 27

Mann A., 2018, Nature, 553, 393

Massol H., et al., 2016, Space Sci. Rev., 205, 153

Matsuhashi H., Tanaka T., Arata K., 2001, J. Phys. Chem. B, 105

Matsui M., 1996, Geophys. Res. Lett., 23, 395

Mazarico E., Neumann G. A., Smith D. E., Zuber M. T., Torrence M. H., 2011, LPI Contributions, 1646, 51

McGovern J. A., Bussey D. B., Greenhagen B. T., Paige D. A., Cahill J. T. S., Spudis P. D., 2013, Icarus, 223, 566

Melosh H. J., 2007, Meteorit. Planet. Sci, 42, 2079

Monaghan J. J., 1992, Annu. Rev. Astron. Astrophys., 30, 543

Monaghan J. J., 2012, Annu. Rev. Fluid Mech., 44, 323

Morbidelli A., Levison H. F., Tsiganis K., Gomes R., 2005, Nature, 435, 462

Morbidelli A., Tsiganis K., Batygin K., Crida A., Gomes R., 2012, Icarus, 219, 737

Morris J. P., 1996, Publ. Astron. Soc. Aust, 13, 97

Nakai Y., Shirai T., Tabata T., Ito R., 1987, At. Data Nucl. Data Tables, 37, 69

Ness N. F., Acuna M. H., Behannon K. W., Burlaga L. F., Connerney J. E. P., Lepping R. P., 1986, Science, 233, 85

Nettelmann N., Helled R., Fortney J. J., Redmer R., 2013, Planet. Space Sci., 77, 143

Nettelmann N., Wang K., Fortney J. J., Hamel S., Yellamilli S., Bethkenhagen M., Redmer R., 2016, Icarus, 275, 107

Nimmo F., Hart S. D., Korycansky D. G., Agnor C. B., 2008, Nature, 453, 1220

Pahlevan K., Stevenson D. J., 2007, Earth Planet. Sci. Lett., 262, 438

Paige D. A., et al., 2014, in LPSC Proceedings. p. 2501

Parisi M. G., Carraro G., Maris M., Brunini A., 2008, Astron. Astrophys., 482, 657

Pearl J. C., Conrath B. J., Hanel R. A., Pirraglia J. A., 1990, Icarus, 84, 12

Petigura E. A., Howard A. W., Marcy G. W., 2013, Proc. Nat. Aca. Sci. USA, 110, 19273

Pierens A., Raymond S. N., Nesvorny D., Morbidelli A., 2014, Astrophys. J. Lett., 795, L11

Podolak M., Helled R., 2012, Astrophys. J. Lett., 759, L32

Poppe A. R., Halekas J. S., Sarantos M., Delory G. T., 2013, J. Geophys. Res. (Planets), 118, 1934

Prettyman T. H., Hagerty J. J., Elphic R. C., Feldman W. C., Lawrence D. J., McKinney G. W., Vaniman D. T., 2006, J. Geophys. Res. (Planets), 111, E12007

Price D. J., 2012, J. Comput. Phys., 231, 759

Raskin C., Owen J. M., 2016, Astrophys. J., 820, 102

Raymond S. N., O'Brien D. P., Morbidelli A., Kaib N. A., 2009, Icarus, 203, 644

Reinhardt C., Stadel J., 2017, Mon. Not. R. Astron. Soc., 467, 4252

Reinhardt C., Chau A., Stadel J., Helled R., 2019, arXiv e-prints, p. arXiv:1907.09809

Saff E. B., Kuijlaars A. B. J., 1997, The Math. Int., 19, 5

Safronov V. S., 1966, Soviet Astron., 9, 987

Sakuraba H., Kurokawa H., Genda H., 2019, Icarus, 317, 48

Schaller M., Gonnet P., Chalk A. B. G., Draper P. W., 2016, Proc. PASC 16 Conf., pp 2:1–2:10

Schlichting H. E., Mukhopadhyay S., 2018, Space Sci. Rev., 214, 34

Schlichting H. E., Sari R., Yalinewich A., 2015, Icarus, 247, 81

Schorghofer N., Aharonson O., 2014, Astrophys. J., 788, 169

Schorghofer N., Taylor G. J., 2007, J. Geophys. Res. (Planets), 112, E02010

Shuvalov V., 2009, Meteorit. Planet. Sci, 44, 1095

Slattery W. L., Benz W., Cameron A. G. W., 1992, Icarus, 99, 167

Smith G. R., Shemansky D. E., Broadfoot A. L., Wallace L., 1978, J. Geophys. Res., 83, 3783

Snellen I. A. G., Brandl B. R., de Kok R. J., Brogi M., Birkby J., Schwarz H., 2014, Nature, 509, 63

Soderlund K. M., Heimpel M. H., King E. M., Aurnou J. M., 2013, Icarus, 224, 97

Springel V., 2005, Mon. Not. R. Astron. Soc., 364, 1105

Springel V., 2010, Annu. Rev. Astron. Astrophys., 48, 391

Stanley S., Bloxham J., 2004, Nature, 428, 151

Stanley S., Bloxham J., 2006, Icarus, 184, 556

Stern S. A., 1999, Rev. Geophys., 37, 453

Stevenson D. J., 1986, in LPSC. pp 1011–1012

Stewart S. T., Seifter A., Obst A. W., 2008, Geophys. Res. Lett., 35, L23203

Tasker E. J., Brunino R., Mitchell N. L., Michielsen D., Hopton S., Pearce F. R., Bryan G. L., Theuns T., 2008, Mon. Not. R. Astron. Soc., 390, 1267

Teodoro L. F. A., Lawrence D. J., Eke V. R., Elphic R. E., Feldman W. C., Maurice S., Siegler M. A., Paige D. A., 2015, preprint, (arXiv:1501.05712)

Tillotson J. H., 1962, General Atomic Report, GA-3216, 141

Tricco T. S., 2019, Mon. Not. R. Astron. Soc., p. 1989

Tsiganis K., Gomes R., Morbidelli A., Levison H. F., 2005, Nature, 435, 459

Tucker J. M., Mukhopadhyay S., 2014, Earth Planet. Sci. Lett., 393, 254

Udry S., et al., 2014, preprint, p. arXiv:1412.1048 (arXiv:1412.1048)

Urey H. C., 1952, The planets, their origin and development, 115, 689

Urey H. C., 1959, J. Geophys. Res., 64, 1721

Vasavada A. R., Bandfield J. L., Greenhagen B. T., Hayne P. O., Siegler M. A., Williams J.-P., Paige D. A., 2012, J. Geophys. Res. (Planets), 117, 0

Wada K., Kokubo E., Makino J., 2006, Astrophys. J., 638, 1180

Wallace D. C., Sidles P. H., Danielson G. C., 1960, J. Appl. Phys., 31, 168

Walsh K. J., Morbidelli A., Raymond S. N., O'Brien D. P., Mandell A. M., 2011, Nature, 475, 206

Wang J., White S. D. M., 2007, Mon. Not. R. Astron. Soc., 380, 93

Waples D. W., Waples J. S., 2004, Nat. Resour. Res., 13, 97

Warren M. S., Salmon J. K., 1993, Proceedings of the 1993 ACM/IEEE conference on Supercomputing, pp 12–21

Warwick J. W., et al., 1986, Science, 233, 102

Watson K., Murray B. C., Brown H., 1961, J. Geophys. Res., 66, 3033

Wetherill G. W., 1976, LPSC Proceedings, 3, 3245

Wieczorek M. A., et al., 2006, Rev. Min. Geochem., 60, 221

Wieler R., Heber V. S., 2003, Space Sci. Rev., 106, 197

Willis J. S., Schaller M., Gonnet P., Bower R. G., Draper P. W., 2018, Adv. Parallel Comp., 32, 507

Woolfson M. M., 2007, Mon. Not. R. Astron. Soc., 376, 1173

Wyatt M. C., Jackson A. P., 2016, Space Sci. Rev., 205, 231

Zahnle K. J., Catling D. C., 2017, Astrophys. J., 843, 122

Zahnle K. J., Gacesa M., Catling D. C., 2019, Geochem. Cosmo. Acta, 244, 56

Zhu M.-H., Ma T., Chang J., Tang Z., Ip W.-H., Xu A., 2011, Sci. China Phys. Mech. Astron., 54, 2083

Zhu M.-H., Chang J., Xie M., Fritz J., Fernandes V. A., Ip W.-H., Ma T., Xu A., 2015, Earth Planet. Sci. Lett., 418, 172